

542 IeValAsnHisAsnThrThrgInValAlaAsnIleThrlleThrgIyAsn 558
1717 AAGATATTGCTACACCGCAATTAAC.....AACAGCTGGAGTAG 1757
559 GluSerIleThrlAlaProSerAsnLysAsnIleAsnLysLeuAspTy 575
1758 CAAAAAGAAATGGCTACAAACGGTGGTTGGGAGAAAGATACAGCA 1807
575 rSerLysGluIleAlaTyAsnGlyTPrPheGlyGluThrAspLysAsnL 592
1808 AAACGACAGGGCGGTCAACCTGTGTTACGACCCCGCGCAAGACCGC 1857
552 yShAsnGlyTyArgLeuAsnLeuIleTyTrLysProThrThrlAspArg 608
1858 ACCCTGCTGTTCCGGCGCAACAAATTAAACGCAACATCAGCAATC 1907
609 ThrLeuLeuSerGlyGlyThrAsnLeuLysGlyAspIleThrlTh 625
1908 AAACGCAACTGTTTTCACCGCGACACACACCGCAGCCTACATC 1957
625 rLysGlyLysLeuPheSerGlyArgProThrProHisAlaTyAsn 642
1958 ATTAAAGACCATGCTGTCGCAAAAAGGGCATTCCTCGCGGGAATC 2007
642 lSleuAspLysArgTrpSerGluMetGluGlyIleProGlnGlyL 658
2008 GTGTGGACAAACGACTGATCAACCGCACATTTAAAGCGAAACTCCA 2057
659 ValTrpAspTyArgTrpIleAsnArgThrPheLysAlaGluAsnPhe 675
2058 AATTAAAGCGGACAGCGGCTGTTCCCGCAATGTTGCCAAAGTGAAG 2107
675 nIleLysGlyLysLeuAlaValAlaSerArgAsnValSerSerIle 692
2108 GCGATGTCATTGAGCAATCAACCGCAACAGATTTTGTGTCGACCG 2157
692 LysAlaThrThrValSerAsnAsnAlaAsnAlaThrPheGlyAlaVal 708
2158 CAT 2160
709 Asn 709
seq_name: p1r2:S61314
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HPI3
C:date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61314
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the Iga gene encoding Iga1 protease in Neisseria
A:reference number: S61314; MIMID:95302961
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1561 <LOW>
A:Cross-references: EMBL:X82474; NID:9732873; PIDN:CAA57857.1; PID:9732874
C:Superfamily: Iga-specific metalloendopeptidase
C:keywords: hydrolase; metalloproteinase
alignment_scores:
Quality: 1615.50 Length: 1753
Ratio: 1.766 Gaps: 55
Percent Similarity: 52.196 Percent Identity: 28.009
alignment_block:
us-09-303-518d-649 x S61314 ..
Align seg 1/1 to: S61314 from: 1 to: 1561

49 AAACGGCGCGCATTCGCTTCGCTGCTTACTAGCATATGCTGTC 98
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheL 18
99 GTTCGGCATTCCTCCCAACGCTGGCGGACACACTTATTTCCGATCA 148
18 aTyAlaLeuThrProTySerGluAlaLeuValArgAspAspValA 35
149 ACTAACCAATCTATCGGACTTTGCCGAATAAAGCAATTTGCAGTC 198
35 sPrTrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
199 GGGCGCAAGATATGTAGGTTTACACAAAAAAGGAGTGGTGGCA 248
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleLys 68
249 ATCAATGACAAAAGCCCGATGATGATTTTCTGTGTCGCGCTAC 297
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsn 85
298 . GCGGTGGCGCATTCGTCGGGATCAATATTTGTGACCGTGCACAT 345
85 yAlaGlyIleAlaThrValAlaAspProGlnTyAlaValSerValLysHis 101
346AACGGCGCTATACACGCTGA 368
102 AlaLysAlaGluValHisThrPheTyTrpGlyGlnTyAsnLysHis 118
369 TTTTGTGGCGGAGAAAGAAATCCGATCAACATCGTTTACTTATMAA 418
118 nAspValAlaAspLysGluAsn.....GluTyArgV 129
419 TGTGCAACGCAATATTTATMAA.....GCAGGAGCT 450
129 aValAlaGluGlnAsnTyGluProHisLysAlaTrpSerIleAsn 145
451 AAAGCGCATTCCTTATGCGCGATATATCAATGCGCGTTGCAATTAAT 500
146 LeuGlyArgLeu.....GluAspTyAsnMetAlaArgPheAsnLysPh 160
501 TGTCAACAGATGCAAGACCTGTTGAATGACCACTTATGATGGG...C 547
160 eValThrGluValAlaProIleAlaProThrAspAlaGlyGlyLeuA 177
548 GGAATATATGATCAAAATTAATTAACCTGACCGTGTGTTGGGCA 597
177 sPrThrTyTrpLysAspLysAsnArgPheSerSerPheValArgAlaGly 193
598 GCGAGCAAA.....TATTGGCGATCTGATGAAGATGACCCCAATTAACCG 641
194 GlyArgGlnLeuValTyTrpGlyLysGlyAlaTyTrpHisGlnGlyLys 209
642 CGAAAGTTCAAT.....CATATTCGAAGTGGATTTGTTGGCTGC 682
210 .GluLysGlyTyAspLeuArgAspLeuSerGlnAlaTyArgTyAlaI 226
683 TTGCTGCG.....AATNC 696
226 IeAlaLysThrProTyTrpLysAspIleAsnIleAspIleThrMetLysThr 242
697 TTTGCACAAATGATCAGTGTGGTGGCAGTCACTTATGATGTAATAA 746
243 GluGlyLeuIleLysPheGlyAsnHisAsnThrHisTySerAlaGlu 259
747 AATTAAACATAGC...CAATATGCTTTTAAACAGAGAGCTCATTTG 793
259 uLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyTrpGlyValLeuG 276
794 GCGAGTGTGCTCACCAATGTTATCTATGATGCCAAAGCAAGCAAGTGG 843
276 LysAspSerGlySerProLeuPheAlaIleAspLysGlnLysAsnGln 292
844 TTAATTAAATGGGTAATTCAAACGGCGCAACCCCTATATAGCAAAAGCA 893

293 ValPheLeuGlyThrTtyAspTyrTrpAlaIlyr...GlyLys..... 306
894 TGGCTTCACAGCTGGTTCGTAAGATGGCTTCATGATGAATCTTTGCTG 943
307LysSerTrpGlnGlyWtrPasnIleYr.... 315
944 GAGATPACCATTCACTATTTCATACGACACAGCTCAAAATGGGAATACTCT 993
316LysLysGln 318
994 TTTAAACAC.....GATATATATGCACAGAGAAAATCA 1028
319 PheAlaAspLysIleLysGlnArgAspAsnIlaGlyThrIleLysGlyTy 335
1029 TGCCAACATGACACAAATCTCTGCTCATATAGATTAATAAACGACGACG 1078
335 rGlyGlnHisThrLysThr.....ThrGlyThr. 345
1079 TTTCAATGTTTAAATGTTCTTTATCCGAGACAGACAGAAACCGTTTAT 1128
346AsnSerHisIleGlySerThr..... 352
1129 CATGCTCAGTGGTGGTCAACAGTTATGACCCAGCTCAATTAATGAGA 1178
353 ...AlaValArgLeuAlaAsnGlnArgSp...AlaAsnGlnGly 367
1179 AAATATTTCTTTATTATGACGAAAGAAAGCGAATTGATCTTACACAGA 1228
367 nAsnValThrPhe.....GluAsnGlnGlyThrLeuValLeuAspGlnA 382
1229 AACTCAATCAAGCTGCTGAGAGATATATTTTCCAAAGAGATTTAAGGTC 1278
382 snIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThrVal 398
1279 TCGCCTGAAAATTAACGA...ACTTGGCAAGCGCGGCGTTCAATACG 1325
399 LysGlyAlaAsnAsnAspIleThrTrpGlnGlyIleAspValAl 415
1326 TGAAGCAGTACCCTTACTTGGAAAGTAAACGGCGGTGCAACACACGCC 1375
415 aAspGlyLysValValThrGlnValLysAsnProAsnLysParG 432
1376 TGGCCAAATTCGCAAAAGCAGCTGCAGCTTCAAGCCAAAGGGGAAAC 1425
432 euAlaLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsn 448
1426 CAAGCCTCAGTACGCGTGGCGACGATCACTATTTGGATCAGCAGCG 1475
1476 AGACGATAAAGCCAAAACCAAGCCTTTAAGTAAATCGCCTTGGTCAAG 1525
465 aAspSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValSerG 482
1526 GCAGGGGTACGGTGCACACTGAATGCCGAATACAGTTCAACCCGAGAAA 1575
482 LysArgIlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsn 498
1576 CTCTATTTCGGCTTTCGCGGCGGACGTTTGAATTTAAACGGGCAATCGCT 1625
499 LeuTyrPheGlyPheArgGlyLysIleArgLeuAspAlaGlnGlyAsnAspLe 515
1626 TTTCGTTCCACCGATTTCAAAATACCGATGAAGGGCGGATGATTCACAC 1675
515 utrrPheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnH 532
1676 ACAATACAAACAAAGATACACCGTTTACCATTTACAGC..... 1713
532 isAsnThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeuIle 548
1713 1713

549	ThrasProAnSerLeuSerValHisSerIleGlnAsnAspTryrAspGI	565
1714ATA	1717
565	uAspAspTryrSerTryrTryrArgProArgProIleProGlnIlyL	582
1718	AGAGATTT.....	1725
582	yAspAspLeuTryrTryrLysAsnTryrArgTryrAlaLeuLysSerGIyLy	598
1726GCTACACCGGCAATAC.....	1743
599	SerValAsnAlaProMetProGlnAsnGIyGlnThrGlnuAsnAsnAspTr	615
1744ACA	1747
615	PIleLeuMetGIySerThrGlnGluAlaLysLysAsnAlaMetAsn	632
1748	GCTTGGATTCGCAAAAGAAATTCCTCAACGGTTCGTTGGCGAGAA	1797
632	IsLysAsnAsnGlnArgIleSerGIyPheSerGIyPhePheGIyGlnGlu	648
1798	GATACGACACCAAAACGACGGCGGCTCAACCTTGGTTCACGCGCCGCGC	1847
649	AsnGIyLysGIyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGIyLysSe	665
1848	AGAGACCGCCACCTGCTTCCTCCGGCGGACAATTTAAACGGCACA	1897
665	AlaGlnAsnArgPheLeuLeuThrGIyGlyThrAsnLeuAsnGIyLysI	682
1898	TCACGCAAAACCAACGGCAACCTGTTTCACGCGCGACACCAACCGGCAC	1947
682	LeSerValThrGlnGlyAsnValLeuLeuSerGIyArgProThrProHis	698
1948	GCCTACATCATTTTAAACACCATTCGTCGCAAAAGAGGC.....AT	1991
699	AlaArgAspPheValAsnLysSerSerAlaArgLysAspAlaArgPheSe	715
1992	TCCTCGCGGGGAATTCGTGGGACACGACTGCATCAACGCACATTTA	2041
715	ILysAsnAsnGluValAlaPheGluAspAspTrIleAsnArgThrPheI	732
2042	AACGGAAAACTTCCAAATTTAAAGCGGACAGCGCGTGTTC...CGC	2088
732	yAlaIleAlaGluIleAlaValAsnGlnSerAlaSerPheSerSerGIyArg	748
2089	AATGTCGCCAAGTGAAGTGAAGCATTCGATTCGACATTCACGCCCAACG	2138
749	AsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysVa	765
2139	AGTTTTGGTGTGCGACCGCATCAAGCCACACAATCTGTACAGCTTCGG	2188
765	AsnLeuGIyTryr...LysAsnGIyAspIleValLysValArgSerA	780
2189	ACTGGACAGGGTCTGCACAATTTGTGTGCAAAAAACCATTCGACAGATTA	2238
780	sPlyrThrGIyTryrValThrCys.....	787
2239	GTCATTGCTTCATTCATCAAGACGACATCAGCGCAATGTCATCTTCGC	2288
788AsnThrAsp.....AsnLeuSe	793
2289	CGATCAACGCTCATTTAATCTCACAGGGCTTCGACACTCAACGGCAATC	2338
793	rAspLysAla.....LeuAsn.....	798
2339	TTTAGTCGCAATGGCGCATACAGTTTACAGTACGCACACGCCACCA	2388
799SerPheAspAlaThrGln	804
2389	..AAGGCACACTTCGCTCGGGGCATGGCCACAGCAACAATTAATCA	2435
805	ILeAsnGIyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGIyLy	821


```
2436 AGCCACATTAAACGGCAACATCCGGCTTCGGCAATGCTTCAATTAAAC 2485
      :|||||  |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
821  sAlaIaLeuTrpGlyGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 838
2486 TAAGCGACCAACGCCGTACAAAAGCGACGTGACGCTTCCGGCAACGCT 2535
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
838  euAsnGlnHis 841
2536 AAGGCAACGTAAGCCATTCCGACGTCAACGCTAATGTCTCCCTAGCCGA 2585
841  ..... 841
2586 TAAGCAGATTATTCATTTTGAAGCAGCGGCTTTACGCGAACAATCAGCG 2635
841  ..... 841
2636 GCGGCAAGGATACGGCATTACCTTAAAGAAGACAGCAATGACGCTGCGG 2685
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
842  .....SerLysTrpHisLeuThr 847
2686 TCAGGACGCGAATTAGCAATTAAACCTTGACAACGCGCACCATTAACACT 2735
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
848  SerAspSerGlnValHisAsnLeuSerIleuAlaAspSerHisIleHisLe 864
2736 CAATTCGCCCTATGCCACAGATGCGGCGGCGCAAAACGCGCAGTGCGA 2785
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
864  uAsnAsnAla.....SerAspAlaGlnSerAla..... 873
2786 CAGATGCGCGCGCGCGGCTTCGCGCGCTTCCGCTTCCTATTATTC 2835
873  ..... 873
2836 GTTACACGCGCAACTTCGCTAGTAATCCGCTTTCACAACGCTGACGCTAA 2885
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
874  .....AsnLysTrpHisThrLeuLysIleAs 882
2886 CGGCAATTTAGCGGTCAAGGACATTCGCTTATGTGCGAATCTTCG 2935
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
882  n...HisLeuSerGlnLysGlnHisIleHisIleHisIleHisIleHisIle 898
2936 GCTACCGGACGACAAATTGAAGTGGCGGAAAGTCCGGAAGCACTTAC 2985
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
898  yAsnLeuGlyAspLysValLeuValLysGlnSerAlaSerGlnHisIleThr 914
2986 ACCTTGGCGGTCAACAATACCGGCAAGCACTGCAAGCTTCGACAAATT 3035
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
915  GlnLeuHisValGlnAspLysThrGlyLysPro...AsnGlnIleGlyLe 930
3036 GACGCTAGTGAAGAAAGCAACAACCGCTGTCCGAAACCTTAATT 3085
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
930  uAsnLeuHisAspAlaSerSerValGlnAspArgSer...ArgLeuSerVal 946
3086 TCACCTCTGCAAAACGACACGCTGATGCGGCGGTGCGGCTTCCCAACGC 3135
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
946  alSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuHisGlyTrpTrpIle 962
3136 ATCCGCAAAAGCGCGAGTTCGCGCTGCAATATCCGCTCAAGAA.... 3180
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
963  LysThrGlnLysGlnGlyIleThrArgLeuLysAsnProTyrAlaGlnLysAsn 979
3180  ..... 3180
979  gATgATgValLysProAlaProSerProAlaThrAsnThrAlaSerGlnA 996
3180  ..... 3180
996  laGlnThrAspSerAlaGlnIleAlaLysProGlnAsnIleValAla 1012
3181  .....CAAGAGCTTTCCGACAAACCTGCG 3203
1013  ProProSerProGlnAlaAsnGlnAlaGlnIleLysArgGlnGlnAla 1029

3204 CAAGCAGAA..... 3213
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1029  alYsAlaGlnGlnValLysArgGlnGlnAlaGlnAlaGlnLysValAlaHis 1046
3213  ..... 3213
1046  isGlnLysAlaGlnLysAlaLysArgGlnGlnAspAlaLeuAlaArgGln 1062
3213  ..... 3213
1063  GlnAlaGlnGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnAlaGln 1079
3214  ....GCCAAAAACAGCGGGAAGAAAGACACGCGCAAGCTTGACGCGC 3259
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1079  uIleAlaLysGlnLysAlaGlnAlaGlnLysArgGlnLysArgGlnAlaGln 1096
3260  TGATTGCGCGCGCGCGATGCGCTGCAAAAGACAGAA...AGCGTTGCC 3306
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1096  LuIleAlaGlnGlnLysAlaAlaAlaGlnLysAlaLysArgGlnAlaGln 1112
3307  GAACGCGCGCGCGAG.....GCAGCGGCGGCAAAATGTCGCG 3341
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1113  GlnLeuAlaArgGlnGlnGlnGlnLysAlaArgLysAlaAlaGlnLysAlaAl 1129
3342  CATTAATGACGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3370
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1129  alYsGlnLysAlaGlnLysArgGlnLysArgGlnLysArgGlnLysArgGln 1146
3371  GGGTCAGCGCGATTAAGACACCGCC...TTGGGCAACAGCGC...GAA 3414
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1146  yAlaGlnAlaGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 1162
3415  GCGGAACCGCGCGCGCTACACCGCGCTCCCGCGCGCGCGCGCGCGCGCG 3464
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1163  GlnGlnGlyArgGlnAlaAlaGlnSerGlnProLysArgGlnArgGlnArg 1179
3465  C.....CGGAGTTTCCGCG 3478
1179  gAlaIleProProGlnLeuSerSerAspAlaThrThrArgAlaLeuProA 1196
3479  AACTG..... 3483
1196  rgIleAlaArgAsnSerAsnProAspAlaSerAspTyrGlnGlnIlePro 1212
3483  ..... 3483
1213  LeuAspAlaLeuGlnAspGlnAspValSerGlnSerValAspThrSerAs 1229
3483  ..... 3483
1229  pLysGlnProGlnAspAsnThrGlnLeuHisGlnLysValGlnThrValS 1246
3484  .....CAACCC 3489
1246  erLeuGlnProArgAlaAlaGlnProArgAlaGlnAlaAlaAlaGlnPro 1262
3490  CAACCGAGCGCCCAACCGCGACGCGCTGATGACCGCTTAAGCCAAATAG 3539
      |||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1263  GlnAlaGlnAlaValAlaGlnAlaAspAlaValAlaSerThrAsnThrAsn 1279
3540  CGGTTTGAGTGAATTTCCGCGACGCTCAACAGCGTTTCCGCG...GTAC 3586
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1279  rAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAspTrpG 1296
3587  AGGACGAATTAGACCGCGTATTTCGCGAAGACCGCGC..... 3624
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1296  lValaSerLeuThrArgHisIleAlaGlnLysSerArgAlaAspAlaGln 1312
3625  ....AACGCGCTTTGGACAGCGCATCCGCGGACACCAAAACCTACCGTTC 3671
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
1313  LysAsnSerValTrpMetSerAsnIleGlyTyrGlyArgAspTyrAlaSer 1329
3672  GCAGAATTTCCGCGCGCTTACCGCGCAACAAACGACCTGCGCAAAATCGGTA 3721
```

```

1      .....|||      :|||      |||||
1329  lrlaglnlyrargarpheSerSerlyargthrclnhrlnlleglyl 1346
3722  tgcagaaaaactcgccagcggg...ccgctgcgcatcctgtttgsgac 3768
1346  lrasprygsrleuSerGluasMetGlnlleglylValleuthrlyr 1362
3769  AACCGGACGGAACACTTTCGACGAGCGATCGCAACTCGGACGGCT 3818
1363  SerAspSerGlnhlsThrPheaspGlnAlaserglyLysanThr...Ph 1378
3819  TGCCACAGCGCGCTTTCGGCAATACGGCATC...GACAGTCTTACA 3865
1378  eValGlnAlaasnleuthrGlyLysThrlyLeuasnAspAlatpTyr 1395
3866  TCGGCATACGCGGCGGCGGCTTTAGACGCGCAGCTTCACAGCGC 3915
1395  alAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuclnhrGln 1411
3916  ATCGAGGCAAAATCCGCGCGCGCTGTCATTCAGCGCATTCAGCGCAG 3965
1412  GlnlyAlaAsnphasnArgthrSerIleGlnThrGlyLeuthrLeucl 1428
3966  ATACCGCGCGCTTTCGGCGGATTCGCAATCGCACATCGCGCAAA 4015
1428  yAsnThrLeuLysIleasnGlnPheGlnIleValProSerAlaGlyIleA 1445
4016  CGCGCTATTTCGTCAAAAGCGGATTCGCTACGCAAAAAGCTCATATC 4065
1445  rGlyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAsnAspSerAl 1461
4066  GCCACCGCGCGCTTCGATTCACCGCTACCGCGCGGCGGCTTAAGCAGA 4115
1462  LysValSerSerMetSerValLysThrLeuThrAlaGlyLeuAspPheal 1478
4116  TTATTCATTCACCGCGCAACATTCATTCACGCTTATTTAGACC 4165
1478  atyArgPheLys...ValGlyAsnleuthrValLysProLeuLeuSerA 1494
4166  TGTCCTAT...ACCGATCGCGCTTCGGCAAAAGTCCGCAACCGCTCAT 4212
1494  lAlaLysrPheAlaAsnTyrGlyLysGlyValAsnValIglYlAsn 1510
4213  ACCGCGCTATTCGCTCAGGATTCGCAAAAAGCGGATTCGCGGATTCG 4262
1511  SerPheAlaTyrLysAlaAspAsnGlnGlnIlyrSerAla.....G1 1525
4263  CGTAACGCGCGAAATCAAGGTTTCACGCTGCTCCACGCTGCGCGC 4312
1525  yAlaAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleT 1542
4313  CCAAGCGCGCGCACTGGAACGCAACACGCGGCGGCGCATCAATTAAGC 4362
1542  hrLysGlyLysGlnLeuGlnLysSerGlyGlnIleLysIleGln 1558
4363  TACCGCTGC 4371
1559  lLeuArgPhe 1561

seq_name: p1r2:A37023

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
N/Alternate names: immunoglobulin A1 proteinase type 1
C/Species: Haemophilus influenzae
C/Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C/Accession: A37023
R/Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Killian, M.
Infect. Immun. 57, 3097-3105, 1989
A/Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophilus influenzae
A/Reference number: A37023; M01D:89379374
A/Status: preliminary

```

A/Molecule type: DNA
 A/Residues: 1-1541 <PDB>
 A/Cross-References: GB:X64357; NID:943560; PDBN:CAA5708.1; PID:943561
 A/Experimental source: serotype b
 C/Superfamily: Iga-specific metalloendopeptidase
 C/Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1608.50 Length: 1717
 Ratio: 1.733 Gaps: 51
 Percent Similarity: 54.048 Percent Identity: 27.257

alignment_block:
 US-09-303-518D-649 x A37023 ..

Align seg 1/1 to: A37023 from: 1 to: 1541

```

64  CGCTTCGCGCTGCTTACTTACCATATGCGTTCGTCGCGCATTCCTCC 113
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5  LysrPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrpr 21
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
114 CCAAGCTGGCGGAGACACTTATTCGCGATTCGCAATACCAATATCATC 163
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
21  otyrthrGlnAlaAlaLeuValArgAspAspValAspTyrIleIlePheA 38
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
164 GCGACTTGGCGGAATTAAGCAAGTTTCGATCGCGGCGGCAAGATATT 213
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
38  rGAspPheAlaGlnAsnLysGlyLysPheSerValIglYAlaThrAsnVal 54
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
214 GAGGTTCACAAAAAAGGAGTGTGCGCAATCATGACAAAGCAAGC 263
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
55  LeuValLysAspLysAsnAsnLysAspLeuclYThrAlaLeuProAsnG1 71
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
264 C...CCGATGATTCATTCCTGTCGTCG...CGTAACGCGCTGCGCG 307
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
71  yLleProMetCileAspPheSerValAlaAspValAspLysArgylealAt 88
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
308 CATTCGTCGCGCATATATATTCGAGCTGCGCAT..... 345
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
88  hrLeuIleAsnProGlnTyrValAlaGlyValLysIleValSerAsnGly 104
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
346 .....AACGCGCGCTTACCAACGCTGA 368
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
105 ValSerGlnLeuIlePheGlyLysLeuAsnGlyAsnMetAsnGlyAs 121
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
369 TTTTGTGCGGAAAGAAATCCGATCAACATCGTTTACTTATTAATAA 418
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 nAlaLysAlaIleAs..ArgAspValSerSerGlnLysAsnArgTyrPheS 137
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
419 TTGTGAACGGAATATTAAGCAGGAGCTAAAGC..... 456
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
137 erValGlnLysAsnGlnIlyrProThrLysLeuAsnGlyLysThrValThr 153
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
457 .....CATCCTTATGCGGCGGATTCATATGCTCCGCGTTC 491
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
154 ThrGlnAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgPhe 170
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
492 GCATTAATTTGTCAGATGACAGAACCTGTGA...ATGACAGATTAA 538
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
170 uAspLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaSerS 187
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
539 TGATGCGCGGAAATATATTCGATCAAAATTAATTAACCTTCGCGTGT 588
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
187 erAspAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArg 203
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
589 ATGCGGCGAGCAGCAATATTCGATGCTGATGAAGATGAG..... 630
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
204 LeuGlySerGlySerGlnPheIleTyrLysGlyLysAsnTyrSerle 220
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
631 .....CCCAATTAACGCGGAAGTTCAATATCAT.....ATTGCA 664
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
220 uIleLeuAsnAsnIleGlnValAlaGlyLysAsnAsnLeuLysLeuValAla 237

```

```

665 GGGGAGATTTCTGGCGTGGTGGGCAATACCTTTGCACAAATGATCA 714
    ::::::::::::::::::::: 111 ::::
237 SPALSTYTHITGELYLLLEAGELYTHProTYrValAsnHISGLY 253
    ::::::::::::::::::::: 111 ::::
715 GGTGGTGGCACAGTCACCTTGGTAGAGAAAATTAAACATGCG...CG 761
    ::::::::::::::::::::: 111 ::::
254 AsnANGlyLeuILEgLYpHeGLYAsnSerLYSGluGlnHISSerSPR 270
    ::::::::::::::::::::: 111 ::::
762 ATATGGTTTTTACCACAGCA.....GGCTCATTTGGCG 796
    ::::::::::::::::::::: 111 ::::
270 OlySLYLeuLeuSerGlnAsPProLeuTHrAsnTYrAlaValLeuGLYA 287
    ::::::::::::::::::::: 111 ::::
797 ACAGTGGCTCACCATGATTATCATGATGCCAACCAAGCAAGTGGTA 846
    ::::::::::::::::::::: 111 ::::
287 SPserGLYSerProLeuPheValTYrAsPArgGLYSGLYSerLYSTrPLeu 303
    ::::::::::::::::::::: 111 ::::
847 ATTAATGGCGTATTGCAACGGGCAACCCCTATATAGAAAAGCAATGG 896
    ::::::::::::::::::::: 111 ::::
304 PheLeuGLYSerTYrAsPThrPAlaGLYTrAsnLYSLYSer..... 318
    ::::::::::::::::::::: 111 ::::
897 CTTCACAGTGGTTCGTAACATTGG...TTCTATGATGAATCTTGGTG 943
    ::::::::::::::::::::: 111 ::::
319 .....TrpGlnIuTrpAsnILETYrLYSerGlnPheThrL 331
    ::::::::::::::::::::: 111 ::::
944 GAGATACCCATTAGTATTCTACGACCAACGTCACAAATGGAATACTGT 993
    ::::::::::::::::::::: 111 ::::
331 yAsPVal..... 333
    ::::::::::::::::::::: 111 ::::
994 TTTAACACAGCATATAATGCGACAGAAAATCATATGCCAACATGACA 1044
    ::::::::::::::::::::: 111 ::::
334 LeuAsnLYSAsPserAlaGLYSer...LeuILEgLYSerLYSThrAsPty 349
    ::::::::::::::::::::: 111 ::::
1044 CAATTCTCTCCCAATAGATTAAACACGCAACCGTT.....C 1081
    ::::::::::::::::::::: 111 ::::
349 rSerTrpSerSerAsnGLY...LYSThrSerThrILEThrGLYGLUL 365
    ::::::::::::::::::::: 111 ::::
1082 AATTGTTTATGTTCTCTTATCCGACACAGCAAGAACCTGTTTATCAT 1131
    ::::::::::::::::::::: 111 ::::
365 ySerLeuAsnValAsPLeuAlaAsPGLYAsPlyPro..... 378
    ::::::::::::::::::::: 111 ::::
1132 GCTGCAGGTGGTGTCAACAGTTATGACACCACAGCATGAATATGAGAAA 1181
    ::::::::::::::::::::: 111 ::::
379 .....AsnHISGLYLYSse 383
    ::::::::::::::::::::: 111 ::::
1182 TATTTCCCTTATGACGAAAGAAAAGCGCAATTGTACTTACCAGCAACA 1231
    ::::::::::::::::::::: 111 ::::
383 rValIThrPhe.....GluGLYSerGLYThrLeuThrLeuAsnAsnNI 398
    ::::::::::::::::::::: 111 ::::
1232 TCATCAAGGTGTGTGAGATATATATTCCAAAGGAGATTTTACGGTC... 1278
    ::::::::::::::::::::: 111 ::::
398 LeuSPeAlaGLYLeuPhePheGluGLYAsPTrGLYValLYS 414
    ::::::::::::::::::::: 111 ::::
1279 TCGCGTGAATAATACGAAACTTGGCAAGCGCGGCGGCTTCATATACAGGA 1328
    ::::::::::::::::::::: 111 ::::
415 GLYThrSerAsPAsnThrTrhTrpLYSerGLYAlaGLYAlaSerValAlaGL 431
    ::::::::::::::::::::: 111 ::::
1329 AGACACTACCGTTACTTGGAAAGTAAACGCGCTGGCAACGACGCGCTGT 1378
    ::::::::::::::::::::: 111 ::::
431 uGLYLYSTrValIThrTrPTrLYSValHISAsnProGlnTYrAsPArgLeuA 448
    ::::::::::::::::::::: 111 ::::
1379 CCAAAATCGCAAAAGCAAGCGTGCACGCTTCAACGCAAAAGGGAACCA 1428
    ::::::::::::::::::::: 111 ::::
448 lAlaSLYLeGLYSerGLYThrLeuILEuGLYGLYThrGLYAsPAsnLYS 464
    ::::::::::::::::::::: 111 ::::
1429 GGTCGATACGCGTGGCGGAGGTACACTATTTTGGATCAGCAGCGAGA 1478
    ::::::::::::::::::::: 111 ::::
465 GLYSerLeuGLYValGLYAsPGLYThrValILEuLYSGlnGlnIThrAs 481
    ::::::::::::::::::::: 111 ::::
1479 CGATTAAGGCAAAAACAAGCTTATAGAAATGCGGTGGTGCAGGGCA 1528
    ::::::::::::::::::::: 111 ::::
481 nGLYSerGLY...GlnHISAlaPheAlaSerValGLYILEValSerLYA 497
    ::::::::::::::::::::: 111 ::::

```

[illegible]


```

3574 GTTTCGGCCGTAAGAGCAATTAAGACGGGTA...TTGCCGACAGACGC 3620
1276 GLYLSALALVALSERGINHISIISESGINLEU1METASNSNGI 1292
3621 CCGCAGACGGCGTTTGACAAAGCGGCATCCGGACACCAACTACCTGTT 3670
1292 YGLITYSANVALTRYPVALSERANTHRSEMETASNLASNTYSERS 1309
3671 CGCAAGATTCCGGCCCTACCGCCCAACAAACGACCTGCCCAATCGGT 3720
1309 ERSEGINTRYATGATGPHSESERLYSERLTHRGINLTHRGINLEUGLY 1325
3721 ATGCAGAAAAACCTCGGCAGCGG...CGCGTCGCATCCTGTTTGCGA 3767
1326 TRPSPRLNTHRISESERASNSVALGINLEUGLYGLVALPHETHTRY 1342
3768 CAACCGGACCGAAACACCTTCGACGACGCGCATCGCAACTCGGACGCG 3817
1342 RVALATGASNSERASNSPHASPLYSALATHRSELYSASNTHT...L 1358
3818 TTGCCACGCGCGCGGTTTTCGGCATACGGCATCGAC...AGGTTCTAC 3864
1358 ENALAGINVALASNPHEIYRSELYSTYTRIALASPSNHSITRYPYR 1374
3865 ATCGGCATCGCGCGCGCGGCGGTTTTCAGACGCGCGCTTCAGACGCG 3914
1375 LEUGLYLEASPLEUGLYTRYGLYLSPHGINSELYSLEUGLTHHS 1391
3915 CATCGGAGGCAAAATCCGGCGCGCGCTGCTGCATACGCGCATTCAGCAC 3964
1391 NHSASNALALYSRPHALATRGHLSHVALAGINPHEGLYLEUTHRALAG 1408
3965 GATACCGCGCGGTTTCGGCGATTCGCGCATCGCAACCGCATCGCGCA 4014
1408 LYSALAPHEASNLISLEUGLYASNPHEGLYIETHPROILEALIGLYVAL 1424
4015 ACCGCGTATTGTCGCAAAACCGGATTCACGCTACGAAACGTCATATAT 4064
1425 AGTGYSETRYLEUSERSNALASPRHEALALEUASPGINALATRGII 1441
4065 CGGCACCGCGCGCGCTTCGATTCACCGCATCGCGCGCGCATTTAAGCAG 4114
1441 ELYSVALASNPROIIESERVALLYSHTHALAPHEALAGINVALASPLEUS 1458
4115 ATTATTCATCAACCGCGCGCACACATTCCTCAGCGCTTATTTAGC 4164
1458 ERTYTHTRYTHRISEUGLYGU...PHESERVALTHPROILEUSER 1473
4165 CTGTCTATACGATCGCGCTTCGCGCAAAAGTCGCAACCGCGCATATAC 4214
1474 ALAATGYT...ASPALASNSINLEIYSEGLYLSIIASNSVALASNGI 1489
4215 CGCGCATTTGCTCAGATTCGCGCAAAACCGGATCGGATCGGAGCGG 4264
1489 YTYRASPRHEALATYRASNSVALIGLASNGINLNLITRYASNLALIGLY 1506
4265 TAAACGCGCAATCAAGGTTTCACGCTGCTCCACGCTGCCGCGCGC 4314
1506 EULYSLEULYSYTHRIASNSVALIYLSLEUSERLEUIEGLYGLYLEUTH 1522
4315 AAAGGCGCGCAACTGGAAGCGCAACAGCGCGCGCATCAATTAAGCTTA 4364
1523 LYSALALYSGINALIGLYLSGLNLYSHTHALALALALALALALALAL 1539
4365 C 4365
1539 e 1539

```

seq_name: p1r2:B41859

seq_documentation_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae

C:Species: Haemophilus influenzae
 A:Variety: strain HK393
 C:date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
 C/Accession: B41859
 R:Poulsen, K.; Reinholdt, J.; Killian, M.
 J. Bacteriol. 174, 2913-2921, 1992
 A:title: A comparative genetic study of serologically distinct Haemophilus influenzae
 A:reference number: A41859; MUID:92234949
 A:Accession: B41859
 A:Status: preliminary; not compared with conceptual translation
 A:molecule type: nucleic acid
 A:Residues: 1-1545 <PDB>
 A:Cross-references: GB:M87490; NID:g148908; PIDN:AAA24967.1; PID:g148909
 A:Experimental source: strain HK393
 A:Note: sequence extracted from NCBI backbone (NCBI:97283)
 C:superfamily: Iga-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality:	1579.50	Length:	1735
Ratio:	1.704	Gaps:	53
Percent Similarity:	53.429	Percent Identity:	26.686

alignment_block:

US-09-303-518d-649 x B41859 ..

Align seg 1/1 to: B41859 from: 1 to: 1545

```

64 CGCTTCGCGCGCTTACTTACCAATATGCGCTGCTGCGCATTTCTCC 113
5 LYSRPHENYSLEUSNPHEIHALALALALALALALALALALALALALAL 11
114 CCAAGCGCTGGCGGCGGACACACTTATTTGCGCATCAACACCAATAC 163
21 OYTRHGINALALALALALALALALALALALALALALALALALALAL 38
164 GCGACTTTCGCGCAAAATTAAGCAAGTTTGCAGTGGGCGGCAAAATTT 213
38 RGRASPRHEALALALALALALALALALALALALALALALALALALAL 54
214 GAGGTTTACCAAAAGGAGGAGTGTGCGGCAATCAATCAATCAATCA 263
55 GLIIVALARGASPLYSASNSANRPROLEUGLYSVALLEUPROASNGI 71
264 C...CGATGATGATTTTTCGTGCTGCTG...CGTAACGCGCTGGCG 307
71 YLLEPROMETILEASPRHESERVALALASPRVALASPRYSARGILEALAT 88
308 CATTTGGGCGCATCATATATTTGAGCGGCGCAAT... 345
88 HLEUVALASNPROIINTYTRVALIGLYVALIYSHISVALSERASNGLY 104
346 .....AACGCGCGCTTAACAACGTTGA 368
105 VALSERGLILEUNISRPHGLYASNLLEUASNLISYNSMETASNSNGIYAS 121
369 TTTTGTGCGGAGAGAAATCCCATCAACATCGTTTACTTATTAATAA 418
121 NALALYSALAHIS...ARGASPRVALSERSEGLUGIUSANRGTLYTRYL 137
419 TTGTGAAAACGGAATATATAT...AACGAGGAGACT 450
137 HVALIGLYASNSGLIUTYRPROTHRIYLSLEUSNGIYLSALALALAL 153
451 .....AAAGGCGCATCTTATGCGCGGATTCATATATGCGCGCTTT 491
154 THRGUASPRGINALALALALALALALALALALALALALALALALAL 170
492 GCATTAATTTGTACAGATGACAGACCTGTGA...ATGACAGGTTATA 538
170 UASPLYSRPHVALTHRGIVALLALALALALALALALALALALALALAL 187

```

```
539 TGGATGGCGGGAATATATCATCAAAATATATACCTGACCGTGTGCT 588
    |||
    |||
187 eethrlaaglythrtyrasnasllyasplstyrprotyrphelvaltyr 203
    |||
    |||
589 ATTGGGCGAGCGCAGCAATAT..... 609
    |||
    |||
204 leuglyserglythrlnphelelytyrgluasnnglythargtygclule 220
    |||
    |||
610 .TGGCATCTGATGAAGATGAGCCCAATAACCGGAAGTTCATATCAT. 657
    |||
    |||
220 utrpheuglylysgluclylnlyserasp..Alaglyglytyrasnl 236
    |||
    |||
658 .....ATTGCAAGTGGCTATCTCTGGCTGCTGGTGGCAATACCTTT 699
    |||
    |||
236 eulysleuvalgllyasnlatytrthtyrglylelaleaglythrprotyr 252
    |||
    |||
700 GCACAAATAGATCAGTGGTGGCAGACATCACTTAGTAGT..... 741
    |||
    |||
253 Gluvalasnhtlsgluasnaspglyleuileglyphcglyasnserasnbs 269
    |||
    |||
742 .....GAAAAATTAACATACCCCATATGCTTTT 772
    |||
    |||
269 nglutyrlleasnprolysglulleuSerLysPro.....L 283
    |||
    |||
773 TACCAACAGAGGCTCATTTGGGAGACGTGGTACCACCAATGTTTATCAT 822
    |||
    |||
283 euthraasnyralaivalleuglyaspseryserproleuphevaltyr 299
    |||
    |||
823 GATGCCCAAAAGCAAAAGTGTATTAATTAATGGGTATTCGAACGGGCAAA 872
    |||
    |||
300 AsparnglyulysglylstrpheuleuglySerTyraspyrtyrpal 316
    |||
    |||
873 CCCCTATATAGCAAAACCAATGCTTCACGTGGTTCGTAAGATTGGT 922
    |||
    |||
316 aglytyr.....AsnlyLysSerTrp 324
    |||
    |||
923 TCTATGATGAATCTTTGCTGAGATACCATTCAGATATCTACGAACCA 972
    |||
    |||
324 lnglutrasnilletyrlyserprogluphealaglyllyserglugln 340
    |||
    |||
973 CGTCAAAATGGG.....AATACTCTTTTAACACGATTA 1007
    |||
    |||
341 TyrserylalyseryleuileglySerlyThrAspyrtyrserTrpser 357
    |||
    |||
1008 TAAAGCCACAGCAAAATCATGCAAAACATGACACATCTCTGCTTA 1057
    |||
    |||
357 rasnngly..... 359
    |||
    |||
1058 ATAGATTAAACACAGCAACGTT.....CAATGTTTATGTT 1095
    |||
    |||
360 .....LysThrserThrIleThrlyglyglulysSerleuasnval 373
    |||
    |||
1096 TCTTTATCCGAGACAGCAAGAACCTTTATCATGCTGCAGGTGTGT 1145
    |||
    |||
374 AspleuAlaAspGlyLysAspLysPro..... 382
    |||
    |||
1146 CAACAGATTTCAGACCCAGACTGAATATATGAGAAATATTTCTTTATG 1195
    |||
    |||
383 .....AsnhtlGlySerValThrPhe... 390
    |||
    |||
1196 ACGAAGAAAGCGCAATGATCTTACCAGACATCAATCAAGGTGCT 1245
    |||
    |||
391 ..gluclyserglythrleuthrleuasnasnhtlaspglnglyala 406
    |||
    |||
1246 GGGAGTTTATTTCCAAAGAGATTTTACGGTC...TCGCTGAAATTA 1292
    |||
    |||
407 glytyrleuphepneuglyllyaspyrtyrgluvallysglythrserasp 423
    |||
    |||
1293 CGAAACTTGCAAGCGCGCGGCTTCATATCAGTAAGACAGTACCGTTA 1342
    |||
    |||
423 nthrtrtrtyrlysglyalaglyvalserValalaglyllyserthrValT 440
    |||
    |||
1343 CTTGGAAGTAAGACGGCTGGCAACGACCGCTGTCCAAATATGGCAAA 1392
    |||
    |||
    |||
440 htrtyrlyvalhtlssnproglintyrasparglyleuAlaLysileglylys 456
    |||
    |||
1393 GGCACCGCTCAGCTTCACCCCAAGCAAAAGCAAGGCTGATCAGCGT 1442
    |||
    |||
457 glythrleuilevalgluaglythrlyaspasnlysglyserleuysva 473
    |||
    |||
1443 GGCAGCGGTACAGTCATTTTGGATCACCAGGACAGCATTAAGGCAAAA 1492
    |||
    |||
473 lgllyaspglythrvalilleuenglyglnclnhrasnlysergly...G 489
    |||
    |||
1493 AACACCCCTTTAGTGAATCGCTGTGTCAGCGGAGGGATACGGTGCAA 1542
    |||
    |||
489 lnhlsalphealaserValgllylevalserglyraserthrleuval 505
    |||
    |||
1543 CTGAATGCCGATATAGTTTCAACCCGCAACACTCTATTTCCGCTTTCG 1592
    |||
    |||
506 leuasnaspaspysglinvalasproasnserIletyrphcglyPheAr 522
    |||
    |||
522 ggllygllyarglyleuaspheuasnglyasnserleuthrphesphstlea 539
    |||
    |||
1643 AAATACCGATGAAGGCGGATGATTTGCAACCAATCAAGCAAGAA 1692
    |||
    |||
539 rgnasnleaspglylalaarglyleuvalasnhtlserthrserlyshs 555
    |||
    |||
1693 TCCACCGTTCCATTTACAGCAATTAAGATTAATGCTTACCAACGGCAATPA 1742
    |||
    |||
556 serthrvalthrlethrlyasp...AsnleuIleThrAspProasnbs 571
    |||
    |||
1743 C..... 1743
    |||
    |||
571 nvalserIletyrtyrvallysproleuagluaspaasnprotyrAlaI 588
    |||
    |||
1743 ..... 1743
    |||
    |||
588 learglnllelystyglytyrglnleutyrrpheasnngluasnarg 604
    |||
    |||
1743 ..... 1743
    |||
    |||
605 ThrtyrtyralaleuylslyaspaIaserlleargserglupheprogl 621
    |||
    |||
1743 ..... 1743
    |||
    |||
621 nasnarglyluserasnbsnsertrpleutyrrmetglythnglulysa 638
    |||
    |||
1744 .....AACAGCTTGATACCAAAAGAAATTT 1770
    |||
    |||
638 laaspaIaglnllyasnaIameIasnhtllesasnngluatgmetasn 654
    |||
    |||
1771 GCCTACACAGGTGTGGTGGAAGAAGATACGACCAAAAGCAAGCGCG 1820
    |||
    |||
655 glyrheasnnglytyrphcglyglnglulgu...glylysasnnglylas 670
    |||
    |||
1821 GCTCAACCTGTTTACAGCCCGCGCAGAAAGACCGCACCTGCTGCTTT 1870
    |||
    |||
670 nleuasnvalthrphelysglyllyssergluglnIasnArgPheleuLeut 687
    |||
    |||
1871 CCGCGGAACAATTAAGCGCAACATCACGCAACAAACGCAAACTG 1920
    |||
    |||
687 hrGlyglyThrAsnleuasnnglylaspleuasnvalglnnglyThrLeu 703
    |||
    |||
1921 TTTTTCACGCGGAGCAACACCGACCGCTACAT..... 1956
    |||
    |||
704 pheleuserglyargprothrphohlsalaragspillealaglyIlese 720
    |||
    |||
1957 .....CATTTAAACGACATTTGGTGCAAAAGAAGGCACTTCTCTGCG 1999
    |||
    |||
720 rserthrlyslaspsershtlsserglu.....AsnA 732
    |||
    |||
2000 GGGAAATCGTGTGGCAACAGACTGGATCAACCCGACATTAAGCGGAA 2049
    |||
    |||
```


52 GlyIaIeHrAspLeuSerValIyLysAsnIySArGdIyGlnAsnIleGlyAs 68
 249 ATCAATGACAAAGCCCGATGATTCATTTCTGTGGTGGCGTAAC 297
 68 nAlaLeuSerAsnValProMetIleAspHeSerValAlaAspAlaAsnL 85
 298 ..GGCGGGGGCATGTGGGGCGGATCAATATATTTGTAGCGTGGCAAT 345
 85 ySArGtIleAlaThrValAlaAspProGlnTyTAlaValSerValIySAs 101
 346AACGGCGCTTAACACAGTTGA 368
 102 AlAlaYsAlaGluValHisThrPheThrTyTArgGlnTyTAsnGlnHisAs 118
 369 TTTTGGTGGGGAAGAGAAATCCCGATCAACATCGTTTACTTATATAA 418
 118 nAspValAlaAspLysGlnAsn.....GluTyTArgV 129
 419 TTGTGAACGAGATATATTAAGCAGGAGCTAAAGCGCATCCTTATAGC 468
 129 aValGlnGlnAsnAsnTyTArgPro.....HisLysAlaTTP 141
 469 GGC.....GATTATCATATGCCGCGTTTGA 494
 142 GlyAlaSerAsnLeuGlyArgLeuGlnAspTyTAsnMetAlaArgPheAs 158
 495 TAAATTTGCACAGATGCAGAACCGTTGGAATGACGACGATATATGATG 544
 158 nLysPheValThrGluValAlaProIleAlaProThrAspAlaGlyGly 175
 545 GG..CGGAATATATCGATCAAAATATTAACCTGACCGTGTCTGAT 591
 175 LysLeuAspThrTyTAspLysAsnArgPheSerSerPheValArgTle 191
 592 GAGCGAGCAGGACAA.....TATTGGCATCGATGATAGATGCCAA 635
 192 GlyAlaGlyArgGlnLeuValTyTGlulysGlyValTyTHisGlnGlu 208
 636 TAACCGGCAAGATTCAAT.....CATATTGCAAGCGCTATATTTCT 676
 208 yAsn...GluLysGlyTyTAspLeuAlaGAspLeuSerGlnAlaTyTArgT 224
 677 GGCTGGTGGTGGC..... 690
 224 yTAlaIleAlaGlyThrProTyTAspLysArLleAsnIleAspGlnThrMet 240
 691 AATACCTTTGCACAAATGATGATCAGACGCGTGGCAGATCACTAGTATG 740
 241 AsnThrGlnGlyLeuIleGlyPheGlyAsnHisAsnLysGlnTyTSerAl 257
 741 TGAATAAATTAAACATAGC...CCATATGGTTTATACCAACAGAGGT 787
 257 aGlnGlnLeuLeuSerGlnAlaLeuSerGlnAspAlaLeuThrAsnTyTArg 274
 788 CATTTGGCAGAGTGGCTCACATGTTATGATGATGCCAATAAGCA 837
 274 aLleuGlyAspSerGlySerProLeuPheAlaPheAspLysGlnIySAsn 290
 838 AAGTGGTAAATTAATGGGATTTGCAACGGGCAACCCCTATATAGGAA 887
 291 GlnTyTArgValPheLeuGlnTyTArgTyTArgAlaGlyTyTArgLys 307
 888 AAGCAATAGGCTTCACAGTGGTTGGTAAAGATTTGTTATATGTAATCT 937
 307 sSerTyTArgGlnTyTArgAsnIleTyTAsnLysGlnPheAlaAspLysIle 323
 938 TTGCTGAGATCAACCATTCAGTATTTCTACGAACACGTCAAATAGGGA 987
 324LysGlnHis..... 326
 988 TACTCTTTAAGCAGATATAATGCGACAGAAAATATCCCAACA 1037
 327AspAsnAlaGlyThrValIySAsnGlnAsnGlyGlnI 338
 1038 TGAACACATTTCTGCGCTATATGATTAATAACAGCAACCGTTCAATTGT 1087
 338 sHisTyTLeuThr.....ThrGlyThr..... 345
 1088 TTAATGTTTCTTATACGACAGACGAAGAAGACCGTTTATCATCTGCA 1137
 346 ..AsnSerHisIleGlySerThrAlaValArgLeuAlaAsnGlnIy 361
 1138 GGTGTGTCAACAGTTATGACCCAGACATCAATTAATGGAATAATTTTC 1187
 362 AspAla.....AsnAsnGlyGlnAsnValTh 370
 1188 CTTTATGACAGAGAAAGCGCAATGATACTTACCAACATCAATC 1237
 370 rPhe.....GluAspAsnGlyThrLeuValLeuAsnGlnAsnIleAsnG 385
 1238 AAGTGTGTGAGATTAATTTTCAACGAGATTTATACGTTCCGTCGCA 1287
 385 InGlyAlaGlyGlyLeuPhePheLysGlyAspTyTThrValLysGlyAla 401
 1288 AATAAGAA...ACTGGCAAGCGCGGCGTTCTATATCATGGAACAG 1334
 402 AsnAsnAspLleThrTyTLeuGlyAlaGlyIleAspValAlaAspGly 418
 1335 TACCGTTACTTGGAAATGAACGGCGTGGCAACGACCGCTGCCAAA 1384
 418 sLysValValThrGlnValLysAsnProAsnGlyAspArgLeuAlaLysI 435
 1385 TCGGCAAGGACACGCTGCAGCTTCAAGCCAAAGGGAAACCAAGCGTC 1434
 435 lGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnGlyGln 451
 1435 ATACCGTGGCGAGCGTACAGTCTTTTGGATGACAGACGACGATA 1484
 452 LeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspAlaAs 468
 1485 AGCAAAAACAAAGCCTTTAGTAATAGCGCTTGATGACGCGAGGGT 1534
 468 rLysLysValGlnAlaArgPheSerGlnValGlyIleValSerGlyArgGly 485
 1535 CGGTGCAATGAATGCGATATACATTCACAAACCCGCAACATCTATTTC 1584
 485 hLeuValLeuAsnSerGlnGlnIleAsnProAspAlaLeuTyTArg 501
 1585 GCGTTTGGCGGCGACGTTTGATTAAGCGGATCGCTTGCGTTGCA 1634
 502 GlyPheArgGlyArgLeuAspAlaAsnGlnLysAsnLeuThrPheG 518
 1635 CGGTATCAAAATACCGATGAAGGGCGGATGATTTGCAACACATCAAG 1684
 518 hHisIleArgAsnValArgGlnGlyAlaArgIleValAsnHisAsnThr 535
 1685 ACAAAAGATTCACCGTTACATTCAGGCAATGAAGTATGTGTCAAC 1734
 535 sPheAlaSerThrIleThrLeuThrGly...LysSerLeuIleThrAsn 550
 1735 GCGATTAACACACCTGGATAGC..... 1758
 551 ProAsnSerLeuSerValHisSerIleGlnAsnAspTyTAspGlnAspAs 567
 1759AAAAAGAA 1768
 567 rTyTSerTyTArgTyTArgProArgArgProIleProGlnGlyLysAspL 584
 1769 TTGCGTAC..... 1776
 584 euTyTArgLysAsnTyTArgTyTArgAlaLeuLysSerGlyArgLeu 600
 1776 1776
 601 AsnAlaProMetProGlnAsnGlyValAlaGlnAsnAsnAspTyTLePhe 617

```

1776 ..... 1776
617 eMetGlyTyrThrIngluLualArgLysAsnAlaMetAsnHisLysA 634
1777 ..... AACGTTGGTTGGCGAGAAAGATACG 1803
634 snAsnArgArgIleGlyAspRheGlyGlyPheRheAspLugLugLysnGly 650
1804 ACCAAAGACAGCGCGGCTCAACCTTGTTTACAGCCCGCGCAGAGA 1853
651 LysGlyHisAsnLysAlaLeuAsnLeuAsnRheAsnGlyLysSerAlaGly 667
1854 CGCGACCCCTGCTGCTTTCGCGGAGAACAAATTTAAACGGCAACATCAGC 1903
667 nLysArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysLysLeu 684
1904 AAACAAACGGCAACGTGTTTTCAGCGGACAGCAACACCGCACCTTAC 1953
684 aLThrIngluLysnValLeuLeuSerGlyArgProThrProHisAlaArg 700
1954 AATCATTTAAAGCAGCATTTGGTCGCAAAAAGAGGC..... ATTCCTCG 1997
701 AspRheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAs 717
1998 CGGGGAATGCTGTGGGACAGCAAGCTGATCAACCGCACATTTAAAGCG 2047
717 nAsnGluValValPheGluAspSerPheLysAsnArgThrPheLysAlaA 734
2048 AAAATTCGCAATTTAAAGCGGACAGCGGCTGCTTC..... CGCATGTT 2094
734 laGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2095 GCCAAATGGAAGCGCATTTGGCATTTGAGCAATCAGCCCAACGACATTT 2144
751 SerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLe 767
2145 TGCTGTGCACCGCATCAAGCCACACAACTCTACAGCTTGGAGCTGA 2194
767 uGlyTyr..... LysAsnGlyAspGluValLysValArgSerAspLys 782
2195 CGGGCTGCACAATTTGTGTGCAAAAACCATTCGCCGACGATTAAGGAT 2244
782 hArgIlyValThrCysAsnThrGlyAsnLeuSer... AspLysAlaLeu 797
2245 GCTTATGACTAAGACGACATCAGCGGCAATGTGCTTCCGATCA 2294
798 AsnSerPheAspAlaThrArgIleAsnGlyAsnValAsnLeu..... 811
2295 CGCTCATTTAAATCTCACAGGCGCTTCCACACTCAACGCCAATCTTAGT 2344
811 ..... 811
2345 CAAATGGCGATACAGCTTATACAGTCAGCCACAGCCACCACAAAGCG 2394
812 ..... AsnGln 813
2395 AACCTTAGCCTCGTGGCAATGCCCAACCAACATTTAATCAGCCACATT 2444
814 AsnAlaIleVal..... LeuGlySalaIle 824
2445 AAACGGCAACATCGCTGCGGCAATGCTTCATTTAATCTAAGCGACG 2494
824 uTrpGlyLysIleGlnGlyIngluLysnSerArgValSerLeuAsnGlnH 841
2495 AGCGCGTACAAAGCGAGTGTGACGCTTCCGCAACGCTAAGCAAC 2544
841 IS..... 841
2545 GTAAGCATTCGCACTCAACGTAATGTCTCCAGCCGATAAGCGAGT 2594
841 ..... 841
2595 ATTCCATTTGAAAGCAGCGCTTACCGGCAATCAGCGGCGCAAG 2644
841 ..... 841
2645 ATACGCATTACACTTAAAGACAGCGAAATGAGACGCTCGCGTACGACG 2694
842 ..... SerLysThrHisLeuThrGlyAspSer 850
2695 GAATTAGGCATTTAACCTTGACAAACGCAACATTAACCTCAATTCGCG 2744
852 GlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla 867
2745 CTATCGCCACAGATCGCGGAGGCGCAACCGCGACAGTCSGACATCGCG 2794
867 a..... SerAspAlaGlnSerAla..... 873
2795 CGCGCGCGGCTTGGCGCGCTTGGCCGCTTCCCTATATCCGTTACACG 2844
873 ..... 873
2845 CCAACTTCGTAGTAATCCGTTTCAACAGCTGACGCTAAGCGCAAT 2894
874 ..... AsnLysThrHisThrIleLysIleAsn... HisLe 884
2895 GAACGTCAGGAGACATTCGCGCTTATGCGGACCTTTCGGCTACCGCA 2944
884 uSerGlyAsnGlyHisPheHisIlyLeuThrAspLeuAlaLysAsnLeuG 901
2945 GCGACAAATTTGAAGCTGCGCGAAAGTTCCGAAGGACATTCACCTTGCG 2994
901 LysAspLysValLeuValLysGlnSerAlaSerGlyHisIlyGlnLeuHis 917
2995 GTCAACAAATTCGCGACAGCAACCTGCAAGCCGCAACATTTAGCGTGT 3044
918 ValGlnAsnLysThrGlyGluPro... AsnGlnGlyLeuAspLeuPh 933
3045 GGAAGGAAAGACAAACAAACCGCTGTCCGAAACCTTAATTCACCTTGC 3094
933 eAspAlaSerSerValGlnAspArgSer... ArgLeuPheValSerLeuA 949
3095 AAAACGACACGTCGATGCGCGCGCGCTGCTTACCAACTCATTCGCGAA 3144
949 laAsnHisIlyValAspLeuGlyAlaLeuArgIlyThrIleLysThrGlu 965
3145 GAGCGAGGTTCCGCGCTGCATTAAT..... CCGGT 3173
966 AsnIlyIleThrArgLeuTyrAsnProTyrAlaGlyAsnGlyArgProVa 982
3174 CAAAGACAAAGCTTTCGACAAACTCGCGCAAGCGAAGAACCAAAAC 3223
982 lLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA 999
3224 AGCGGGAAGAAAGACAAACGCGCAACGCTTGACGCGCTGATTGCGCGG 3273
999 laThrGlnThrAspGlyAlaGln..... IleAlaLysPro 1010
3274 CGCGATCGCGCTGCAAAAGACAGAAAGCGTTCCGCAACCGCGC..... 3315
1011 GlnAsnIleVal..... ValAlaProProSerProGlnAl 1022
3316 ..... CGCGACGCAAGCGGCGGGAATGTCG 3340
1022 aAsnGlnAlaGluLugLualLeuArgGlnGlnAlaLysAlaGluGlnAl 1039
3341 GCATTATCCAGGCG..... GAGGAA 3360
1039 ysArgGlnGlnAlaIlaGluAlaGluLysValAlaArgGlnLysAspGlu 1055
3361 GAGAAAGAAAGCGTGCAGCGGAT..... 3384
1056 GluAlaLysArgLysAlaIlaGluIleAlaArgGlnGlnGluLualAar 1072
3385 ..... AAG 3388

```

```

1072 gTysAlaAlaGluLeuAlaAlaLysGlnLysAlaGluAlaGluArgLysA 1089
3389 ACACGCCCTTGCGAAGACGCGCAAGCGGCAACCGCGCGGCTACACAC 3438
1089 laArgGluLeuAlaArgGlnLysAlaGluAlaSerHisGlnAlaAsn 1105
3439 GCGTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGCA..... 3480
1106 AlaLysProLysArgArgArgArgAlaAlaLeuProArgProProAl 1122
3480 ..... 3480
1122 aProValPheSerLeuAspAspLysAlaLysAspAsnSerGluSer 1139
3480 ..... 3480
1139 eRIleGlyAsnLeuAlaArgValIleProArgMetGlyArgGluLeuIle 1155
3480 ..... 3480
1156 AsnAspLysLugLugIleProLeuGluGluLeuGluAspGluAlaGlu 1172
3480 ..... 3480
1172 uGluArgArgGluAlaThrGlnPheHisSerLysSerArgAsnArgAla 1189
3480 ..... 3480
1189 laIleSerSerGluProSerSerAspGluAspAlaSerGluSerValSer 1205
3480 ..... 3480
1206 ThrSerAspLysHisProGlnAspAsnThrGluLeuHisGluLysValG 1222
3481 ..... 3481
1222 uThrAlaGlyLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAla 1239
3506 CGCAGCGCGACCGATCAGCCGTTATGCCAATAGCGGTTTGAGTAATT 3555
1239 laGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAspAla 1255
3556 TCCGCGACCGTCAACAGCGTT.....TTCGCGGTCA 3587
1256 MetaIleSerThrGlnSerIleLeuLeuAspThrGlyAlaTyrLeuThr 1272
3588 GAGCAGATTAAGACCGCGTATTTGCCGAAGACCGCGACCGCTTGGA 3637
1272 gHisIleAlaGlnLysSerArgAlaAspAlaGluLysAsnSerValTrp 1289
3638 CAAGCGGATCCGCGGACCAACACATCCGTTGCGAAGATTTCCGCGC 3687
1289 eIleSerAsnThrGlyTyrGlyArgAspLysAlaSerAlaGlnTyrArg 1305
3688 TACCCGCAACAACCGACCTCGCCCAATCGGTATGCAAAAAACCTCG 3737
1306 PheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgSerLeu 1322
3738 CAGCGGG...GCGCGGCGATCCCTTTTCCGCAACCGGACCGCAACA 3784
1322 rGluAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSerGlnHis 1339
3785 CCTTGACGACGCGATCGCAACTCGGACGCGCTTGGCCCGCGCGCTT 3834
1339 hrPheAspGlnAla...GlyGlyLysAsnThrPheValGlnAlaAsn 1354
3835 TTCGGGCAATACGGCATC...GACAGTTCTACATCGGATCAGCGCGG 3881
1355 TyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyAspIle 1371
3882 CGCGGGTTTGTGACGCGACCGCTTTCAGACGCGATCGGAGGAAATCC 3931
1371 YAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPh 1388
3932 GCCCGCGGTGCTGATTAACGCGATTCAGGACAGATACCGCGCGGTTTC 3981
1388 snArgThrSerIleGlnThrGlyLeuThrLeuGlnLysAsnThrLeuLys 1404
3982 GCGGATTTCCGATCGAACCGCACATCGCGCAACCGCGCTATTTCGTCA 4031
1405 AsnGlnPheGluIleValProSerAlaGlyIleArgTyrSerArgLeu 1421
4032 AAACCGGATTCAGCGCTACGAAACGTCATATCGCACCGCGCGCTTG 4081
1421 rSerAlaAspLysLysLeuGlnLysAspSerValLysValSerMetAl 1438
4082 CATTCACCGCTACCGCGCGGATTAAGCAGATTATTCATTCACACCG 4131
1438 laValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLys 1453
4132 GCGCAACACATTCATCCATCAGCGCTTATTTGAGCGCTGCTATACGAT 4181
1454 ValGlyLysLeuThrValLysProLeuLeuSerAlaAlaTyrPhe... 1469
4182 CGCTTCGCGCAAGTCCGACACGCGCTCAATACCGCGCTATTCGCGCAG 4231
1469 aAsnTyrGlyLysGlyValAlaAsnValGlyLysSerPheAlaTyrL 1486
4232 ATTTCGCAAAACCGCGCATGCGGAAATGGCGCTAAACCGCAATCAA 4281
1486 yAlaAspAsnGlnGlnGlnTyrSerAlaGlyValAlaLeuLeuThr 1502
4282 GGTTCACGCTGTCCTCCACGCGTGGCGCAAGACCGCGCACCTGGA 4331
1503 AsnValThrLeuAsnValAsnLysSerIleThrLysGlyLysGlnLeu 1519
4332 AGCGCAACACAGCGCGCATCAATTAAGCTACCGCTCG 4371
1519 uLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1532
seq_name: p1r2:H64106
seq_documentation_block:
  IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (str
  N:Alterate names: Immunoglobulin A1 proteinase type 1
  C:Species: Haemophilus influenzae
  A:Variety: strain Rd KW20
  C:date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
  R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
  ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidm
  Science 269, 496-512, 1995
  A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent
  A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
  A:Reference number: A64000; MUID:95350630
  A:Accession: H64106
  A:Status: nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-1694 <TIGR>
  A:Cross-references: GB:U32779; GB:I42023; MID:g1574009; PIDN:AAC22651.1; PID:g1574019
  A:Experimental source: strain Rd KW20
  R:Grundy, F.J.; Plant, A.G.; Wright, A.
  Infect. Immun. 58, 320-331, 1990
  A:Title: Localization of the cleavage site specificity determinant of Haemophilus int
  A:Reference number: A41500; MUID:90129281
  A:Accession: A41500
  A:Molecule type: DNA
  A:Residues: 1-377 <GRD>
  A:Cross-references: GB:X59800
  A:Experimental source: strain Rd KW20
  A:Note: the authors translated the codon TGG for residue 319 as Thr
  A:Function:
  C:description: this proteinase is classified as type 1 because it cleaves at a prolin
  C:superfamily: IGA-specific metalloendopeptidase
  C:keywords: hydrolase; metalloproteinase

```


2304	AAATTCACAGGCGCTTCCACACACTCAACGGCAATCTTATGCAAAATGGCG	2353
828	828
2354	ATACACGTTATACAGTACGCCACACGCCAACCCAAAACGGCAACCTTAGC	2403
828	828
2404	CTCGGGCAATGGCCCAAGCAACATTTAATCAAGCCACACTTAAACGCCAA	2453
829Glysnalaaasnphenylleugilylsalaaasleuphegilyth	843
2454	CACATCGGCTTGGGGCAAGCTTCATTATATGTGAAGCCACCGCCGTAC	2503
843rilesergilythrglyasnserglnvalargyleuthrglu	856
2504	AAAAGGCACATCTGACGCTTTCGGGACACGCTAAGGCAAAACGTAGCCAT	2553
856	856
2554	TCCGCACTCAACGTAATGCTTCCTAGCCGTAAGCAAGTATTCATT	2603
856	856
2604	TGAAGCAGCGCTTACCGGCAAAATCAGCGCGGCAAGGATACGCAAT	2653
856	856
2654	TACACTTAAAGACAGCGAATGAGCGCGCTGACGACGACGAATTAGGC	2703
857Asnserhistrphlsleuthrglyasphenasvalasn	869
2704	AATTTAAACCTTGACACGCGCACCATTTACACTCAATTCGCGCTATGCCA	2753
870Glnleuasnleuasprysgilystihleuasnlaa	884
2754	CGATGGGCGACGGGGCGCAAAACCGGCAAGTGCACAGATGCGCGCGCGCC	2803
884naspalaaasnlysvallrthr	891
2804	GTTTCGCGCGTTCGGCGCGTTCCTATTATTCCTTACACGCCAACCTTG	2853
891	891
2854	GTAGATCCCGTTTCAACACGCTACGCGTAAACGGCAATTGAACGTC	2903
892TyrasnhrleuthValasn	903
2904	GGGAACATTCGCTTATGTGTGGAACCTTCCTGCTACCGGACGACAAT	2953
903nglyserphenylrlyleuthhraspleuserasnlysginglyaspyslv	920
2954	TGACGTGGCGGAAAGTTCCGACAGCACTTACACCTTGGCGGTCAACAT	3003
920alvalvalthrlyseralathrglyasnphenrhrleuglnvalalaaasp	936
3004	ACCGCAACGAAGCGGACGCTGCGACCAATTTACAGTAAGCTGAGAGAA	3053
937LysThrnglyluprthrLys	951
3054	AGACAAACAACCGCTGTCCGAAAACCTTAATTTACCCCTGCAAAAGAAC	3103
952SerasnlaThrtrargnaasnlleuasnlvalserleuValgIysAsr	967
3104	ACGTGATGAGCGGGCGCGCGCTTACCACTCATCCGCAAGACGGCGAG	3153
967hvalaspleugilyalatrplysryllysleuarargasnValasnlyArg	983
3154	TTCCGCGCTGCAATATCCG	3171
984	TyraspleuThyrasnProgluValgIlylsArgAsnGlnThrValasPth	1000

```
3171 ..... 3171
1000 rThrasnleThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 ..... 3171
1017 roSerAsnAsnGlnGluIleAlaArgValGluThrProValProProPro 1033
3171 ..... 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 ..... 3171
1050 sGlnGlnSerLyThrValGluLysAsnGlnGlnAspAlaThrGluThrT 1067
3171 ..... 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 ..... 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100
3172 ..... GTCAAAGA... 3180
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluLysAla 1117
3181 ..... CAAGAG 3186
1117 LaLysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSerGlu 1133
3187 CTTTCGCAACTCGGCAG... 3207
1134 ThrSerProLysGlnAlaLysProAlaProLysGluValSerThrAspTh 1150
3207 ..... 3207
1150 rLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSerThrT 1167
3208 ..... GCAGAGCC 3216
1167 hrValAlaAlaGlnAlaThrSerProAsnSerLysProAlaGluGlu 1183
3217 AAAAAAGCGCGAAAGACACACCGCCAGACGCTTGACGCGTATGTC 3266
1184 ThrGlnProSerGluLysThrAsnAlaGluProValThrProValValSe 1200
3267 G.....G 3268
1200 rLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGluLysT 1217
3269 CCGGCGCGGATGCGCTGAAAGACACAGAACGTTGCCGAACCGCGCGG 3318
1217 hrAlaLysValGluThrGluLysThrGlnGluProGlnValAlaSer 1233
3319 CAGGAGGC..... 3327
1234 GlnAlaSerProLysGlnGlnGlnSerGluThrValGlnProGlnAlaVa 1250
3328 .....GGGAAATGTC..GCCATTATCGACGCGAGAGAGAGAAA 3367
1250 lLeuGlnSerGluAsnValProThrValAsnAsnAlaGluGluValGlnA 1267
3368 AACGGGTGACGGGATTAAGAC..... 3390
1267 laGlnLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnProAla 1283
3391 .....ACCGCTTGCGGAAACACGC 3410
1284 ProGluAsnSerIleAsnThrGlySerAlaThrAlaIleThrGluThrAl 1300
3411 CGAACGGAAACCGCGCGGCTACC..... 3435

1111 ... 1111
1300 aGluLysSerAspLysProGlnThrGluThrAlaAlaSerThrGluAspA 1317
3435 ..... 3435
1317 laSerGlnIleLysAlaAsnThrValAlaAlaAspAsnSerValAlaAsnAsn 1333
3436 .....ACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGA 3479
1334 SerGlnSerSerAspProLysSerArgArg...ArgArgSerIleSerG 1349
3480 ACTGCAA..... 3486
1349 nProGlnGluThrSerAlaGluGluThrThrAlaAlaSerThrAspGluT 1366
3487 .....CCCGACCGCGACGCCCAACCGCAG 3510
1366 hrThrIleAlaAspAsnSerLysArgSerLysProAsnAlaArgSerArg 1382
3510 ..... 3510
1383 ArgSerValArgSerGluProThrValThrAsnGlySerAspArgSerTh 1399
3511 .....CGGACCTGATCAGCCGTTATGCCAATAGGGTTGAGTG 3550
1399 rValAlaLeuAlaArgAspLeuThrSerThrAsnThrAsnAlaAlaIleSerA 1416
3551 AATTTCGCCG...ACGTCACACCGCTTTCGCCGTA 3585
1416 sPalaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLysAlaValA 1432
3586 CAGACGAATTAGACCGGCTA...TTTGCCGAACACCGCGCGCGCT 3632
1433 SerGlnIleIleSerGlnLeuGluMetAsnAsnGlnGluGlnTyAsnVa 1449
3633 TTGGACACGGCGCATCCGGACACCAACACACACGTTCCGCAAGATTTC 3682
1449 lThrValSerAsnThrSerMetAsnGluAsnTySerSerSerGlnTyA 1466
3683 CGGCTACCGCCCAACACACGACGCTGCGCAATCGTATGACAGAAAAC 3722
1466 rGArgPheSerSerLysSerThrGlnThrGlnLeuGluTyTrAspAlnThr 1482
3733 CTCGCGACGGG...CGGTCGGCATCCGTTTCCGCAACCGGACGA 3779
1483 lIleSerAsnAsnValGlnLeuGluGluAlaPheTyrTyValAlaArgAsnSe 1499
3780 AACAACCTTCGACGACGCGATCGGACACTCGGACGCGCTTGCCGACGCG 3829
1499 rAsnAsnPheAspLysAlaSerSerLysAsnThr...LeuAlaGlnValA 1515
3830 CCGTTTTCGGCAATACGCGATCGAC...AGGTTCTACATCGGCTACG 3876
1515 snPheTySerLysTyTyAlaAspAsnIleStrTyLeuGluIleAsp 1531
3877 GCGGCGCGGCGTTTACGACGCGACGCTTCGACGCGATCGGACGA 3926
1532 LeuGluTyGluLysPheGlnSerAsnLeuGluThrAsnIleAsnAlaLy 1548
3927 AATCCGCGCGCGCTGCTGATTCAGCATTAAGCAGACATCGCGCGCG 3976
1548 sPheAlaGlnIleThrAlaGlnPheGluLeuThrAlaGluLysAlaPheA 1565
3977 GTTTCGCGCGATTCGGCATCGAAGACGCAATGCGGCGACGCGCTATTC 4026
1565 snLeuGluAsnPheGluTyIleThrProIleValGluAlaArgTySerTy 1581
4027 GTCGAAAGCGGATTCAGCTACGAAAGCTGCAATATTCGCAACCGCGCG 4076
1582 LeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLysValAsnPr 1598
4077 CCGTCATTCACCGGCTACCGCGCGGCGCATTAAGGACGATTTGCTTCA 4126
..... 4126
```



```

1598 oIleSeVallysthrAlaphealagInValAspLeuSerTyrThrTyr 1615
4127 AACGGGCGACACATTTTCATCAGCCTATTGAGCCTGCTCT... 4173
      ::::: :::::||||| |||||
1615 IsteIuglyGlu...PheSerValThrProIleuSerAlaArgTyrAsp 1630
4174 ACCGATCCCGCTTCGGGCAAGTCGACACAGCGGTCATAATGACGCTATT 4223
      ::::: :::::|||||
1631 ThrAsnGlnGlySerGlyLysIle.....AsnValAsnGlnTyrAsp 1645
4224 GGCTCAGATTTTCGGCAAAACCCGACATGGCGATGGCGGTAAAGCGCG 4273
      ::::: :::::|||||
1645 eAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLysLeu 1662
4274 AAATCAAGGTTTCAGCTGCTCCACGCTCCGCGCCGCAAGGCGCG 4323
      ::::: :::::|||||
1662 ystYrHisAsnValLysLeuSerLeuIleGlyLeuThrLysAlaLys 1678
4324 CAACTGAGACGCGACACACAGCGCGGACATCAATTAAGGCTAC 4365
1679 GluAlaGlnLysGlnLysThrAlaGlnLeuLysLeuSerPhe 1692
seq_name: p1r2:A41859

```

```

seq_documentation_block:
  IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
  C:Species: Haemophilus influenzae
  A:Variety: strain HK715
  C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
  A:Accession: A41859
  R:Poulsen, K.; Reinholdt, J.; Kilian, M.
  A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
  A:Reference number: A41859; MUID:92234949
  A:Accession: A41859
  A:Status: preliminary; not compared with conceptual translation
  A:Molecule type: nucleic acid
  A:Residues: 1-1702 <POU>
  A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907
  A:Experimental source: strain HK715
  A:Note: Sequence extracted from NCBI backbone (NCBIP:97282)
  C:Superfamily: IGA-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
  Quality: 1566.00      Length: 1872
  Ratio: 1.671         Gaps: 49
  Percent Similarity: 50.053      Percent Identity: 24.947

```

alignment_block:

us-09-303-518d-649 x A41859 ..

Align seg 1/1 to: A41859 from: 1 to: 1702

```

64 CGCTTCGCGCTGCTTACTAGCATATGCTGCTGCTTGGCATTTCTTC 113
      ::::: :::::|||||
5  LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAGCCTGGCGGCGACACACTTATTCGGCATCACTACCAATATCATC 163
      ::::: :::::|||||
21  otYrThrGlnAlaLeuValArgAspAspValAspTyrGlnIlePhe 38
164 GCGACTTGGCGGAATAATAGCGAAGTTTGACGCGGCGGGAAGATATT 213
      ::::: :::::|||||
38  rGAspPheAlaGlnAsnLysSelYArPheSerValGlyAlaThrAsnVal 54
214 GAGGTTTCAACAACAAAGGAGACTGTGCGCAATCAATGACAAAGC 263
      ::::: :::::|||||
55  GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsn 71
264 C...CGATGATGATTTTCTGCTGCTGCG...CGTACGCGGCGGCG 307
      ::::: :::::|||||

```

```

71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAla 88
308 CATGGTGGGCGAATCAATATATGTGACGGTGGCAAT...AAGCGCGC 354
      ::::: :::::|||||
88  hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATACACAGTTGATTTTGGT.....GCGGAGGAGGAANA 389
      ::::: :::::|||||
105  ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 121
390 TCCGATCAACATCGTTTACT.....TATAAATTG 421
121  AspLysSerHisArgAspValSerSerGlnGlnAsnArgTyrPheSer 138
422 TGAACGGAATTAATTAATTAAGCAGGACATAAGC..... 456
      ::::: :::::|||||
138  alGluLysAsnGlnTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
457 .....CATCCTTAGCGCGGATTAATATGACCGCGTTTGA 494
155  GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeu 171
495 TAAATTTGCACAGATGCAGAACCTGTTGAA...ATGACCACTTATATG 541
      ::::: :::::|||||
171  pLysPheValThrGluValAlaProIleGlnLysSerThrAlaSerSera 188
542 ATGGCGGGAATATATATGATCAAAATTAATTAACCTGACCGTTCGTAT 591
      ::::: :::::|||||
188  spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheAlaArgLeu 204
592 GGGGAGGAGGAGCAATATGCGATCTGATGAAGTGAAG..... 630
      ::::: :::::|||||
205  GlySerGlySerGlnPheIleTyrLysGlyAspAsnTyrSerLeuI 221
631 .CCCAATACCGCGGAAGTTTCATATCAT.....ATTGCAAGTG 667
      ::::: :::::|||||
221  eLeuAsnAsnHisGluValGlyLysAsnAsnLeuLysLeuValGlyAsp 238
668 CGATTTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
      ::::: :::::|||||
238  lArYrThrTyrGlnLysLeuAlaGlyThrProTyrLysValAsnHisGln 254
718 GGTGGCAGATCACTAGTAGTGAATAAATAATTAACATAGC...CGATA 764
      ::::: :::::|||||
255  AsnGlyLeuIleGlyPheGlyAsnSerLysGlnGlnHisSerAspPro 271
765 TGGTTTTTACCACACAGA.....GGCTCATTTTGGCGACA 799
      ::::: :::::|||||
271  sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAsp 288
800 GTGGCTCAACAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAAT 849
      ::::: :::::|||||
288  erGlySerProLeuPheValTyrAsnArgGlnLysGlyLysTyrLeuPhe 304
850 AATGGGATATTGCAACCGGACACCCCTATATAGGAATAAGCATGAGCT 899
      ::::: :::::|||||
305  LeuGlySerTyrAspPheThrAlaGlyTyrAsnLysLysSerThrPhe 321
900 CCAGCTGTTGCTAAAGATTGCTTATGATGAATTTTGTGGAGATA 949
      ::::: :::::|||||
321  uTPAsnLleTyrLysProGlnPheAlaLysThrValLeuAspLysAsp 338
950 CCATTCAGTATCTAGAACACAGTCGCAAAATAGGGAATCTGTTTAAAC 999
      ::::: :::::|||||
338  hrAlaGlySerLeu.....ThrGlySerAsnThrGlnTyrAsnThrAsn 352
1000 GACATATATATGCGACAGGAAAAATCATATGCGAAACATGACACAAATTC 1049
      ::::: :::::|||||
353  Pro.....ThrGlyLysThrSerVal..... 359
1050 TCTGCTAATATAGATTAAACACGACCGCTCAATGCTTAAATGTTTCT 1099
      ::::: :::::|||||
360  IleSerAsnGlySerGlnSerLeuAsnValAspLeuPheAspSerSerg 376

```

```

1100 TATCCGACAGCAGAGAACTGTTATCATGCTGCAGGTGTCTCAAC 1149
    : : : : :
376 lAspThrAspSerLys..... 381
1150 AGTTATGCACCCAGACTGAATAATGAGAAATATTCCTTTATGACGA 1199
    : : : : :
382 .....LysAsnAsnHisGlyLysSerValThrLeu.....Ar 392
1200 AGGAAAGGCGAATGTACTCTTACAGCAACATCAATCAAGGTCTGAG 1249
    : : : : :
392 gGlySerGlyThrLeuThrLeuAsnAsnHisLeuAspGlnGlyAlaGly 409
1250 GATTATATTTTCCAAGAGATTTTACGGTC...TCGCTGAAATAACGAA 1296
    : : : : :
409 LysLeuPhePheGlnGlyAspTyrGlnValLysGlyThrSerAspSerThr 425
1297 ACTTTGGCAAGGCGCGCTTCATATCAGTGAAGACAGTACCGTTACTTG 1346
    : : : : :
426 ThrTyrLysGlyAlaGlyAlaSerValAlaAspGlyLysThrValThrTr 442
1347 GAAAGTAAAGCGCGTGGCAACGACGCGCTGTCCAAATCGCAAGGCA 1396
    : : : : :
442 pLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLysGlyTr 459
1397 CGGTGCAGCTTCAAGCCAAAGGGAACCAAGGCTCGATCAGCGTGGGC 1446
    : : : : :
459 hTrLeuIleValIleGlnGlyLysGlnLysSerLysLeuLysValGly 475
1447 GAGGTACAGTCAATTTGGATCAGCAGCAGACGATTAAGCAAAAAACA 1496
    : : : : :
476 AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysValLys 492
1497 AGCCTTAGTGAATTCGGCTTGGTGCACGGCAGGGGTAGCGTGCACCTGA 1546
    : : : : :
492 sAlaPheSerGlnValIleValIleValSerGlyArgSerThrValValLeuA 509
1547 ATGCCGATTAATCAGTTCACACCCGCAACAACCTTTCCGGTTTCCGGGC 1596
    : : : : :
509 snAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 525
1597 GGAGCTTGGATTTAAACGGGCGATTCGCTTCCGTTCCACCGTATTCAAA 1646
    : : : : :
526 GlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIleArgAs 542
1647 TACCGATGAAGGCGGATGATTTGCACACCAATCAAGACAAAGAATCCA 1696
    : : : : :
542 nIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThrSerT 559
1697 CCGTTACCATTAACAGC..... 1713
    : : : : :
559 hTrValThrIleThrGlyLeuSerLeuIleThrAspProAsnThrIleThr 575
1713 ..... 1713
576 ProTyrAsnIleAspAlaProAspLysAspAsnProTyrAlaPheArgAr 592
1713 ..... 1713
592 gIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGlnAsnTyrThrTyrT 609
1713 ..... 1713
609 yTrAlaLeuArgLysGlyAlaSerThrArgSerGlnLeuProLysAsnSer 625
1714 .....AATAAAGATATTGC 1727
    : : : : :
626 gIyGlnSerAsnGlnAsnTyrPheTyrMetGlyLysThrSerAspGlnAl 642
1728 TACAACCGGCAATTAACACAGCTTGATGACAAAGAAATATGGCTACA 1777
    : : : : :
642 aLysArgAsnValMetAsnHisIleAsnAsnGlnLysMetAsnGlyPheA 659

```

```

1778 ACGGTGTTGTTGGCGAGAAAGATACGACCAAAAGCAAGGCGGCTTCAC 1827
    : : : : :
659 snGlyTyrPheGlyGlnGlnLysLysAsnAsnGlnLysLeuAsn 674
1828 CTGTGTTTACGAGCCCGCGCAGAGACCGCAACCTGCTGCTTCGCGCG 1877
    : : : : :
675 ValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuThrGlyGly 691
1878 AACAATTTTAAAGCGCAACATCAGCAACAAACGAGCAAACTGTTTTC 1927
    : : : : :
691 yThrAsnLeuAsnGlnLysPheLysValGlnLysGlyThrLeuPheLeuS 708
1928 GCGCGACAGCAACACCGCAGCCTTACAT..... 1956
    : : : : :
708 erGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThr 724
1957 CATTTAAACGACCATTTGCTCCAAAAGAGGCAATTCCTCGCGGGGAAT 2006
    : : : : :
725 LysLysAspGlnHisPheAlaGln.....AsnAsnGlnVal 736
2007 CGTGTGGGACACGACGTGATCAACCGCACATTTAAACGGGAAACTTCC 2056
    : : : : :
736 lValValGlnAspAspTyrPheAsnArgAsnPheLysAlaThrAsnIleA 753
2057 AATTAAAGCGCGACAGCGCGTGTTC...CGCAATGTTCCAAAGTG 2103
    : : : : :
753 snValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIle 769
2104 AAAGCGATGTGCATTTAGCAATCAACGCCCAAGAGTTTGTGGTGGC 2153
    : : : : :
770 ThrSerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr.. 785
2154 ACCGCATCAAGCCACACAAATCTGTACACGTTGAGCTGACGGGCTGCA 2203
    : : : : :
786 ...LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 801
2204 CAATTTGTGTCAAAAAACCATTAACGAGCAATTAAGTGTTCCTCATTTG 2253
    : : : : :
801 aLThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 816
2254 ACTTAAGCAGCATCAGCGCAATGTGATCTTGGCGATCAACGCTCATTT 2303
    : : : : :
817 AsnAlaThrAsnValSerIleLysAsnValAsnLeuSer..... 828
2304 AATTCACAGGCGCTTCCACACTCAACGCAATCTTAAGCAATGGCG 2353
    : : : : :
828 ..... 828
2354 ATACAGTTATACAGTCAGCACACCAACGCCAAACGCAACCTTAGC 2403
    : : : : :
828 ..... 828
2404 CTCGTGGCAATGCCACAGCAACATTTAATCAAGCCACATTAAGGCA 2453
    : : : : :
829 .....GlyAsnAlaAsnPheValIleGlyLysAlaAsnLeuPheGlyTh 843
2454 CACATGGCTTGGGCAATGCTTCATTTAATCAAGCAACGACGCGTAC 2503
    : : : : :
843 rIleSerGlyThrGlyAsnSerGlnValArgLeuThrGln..... 856
2504 AAAAGCGACTGTGAGCTTTCCGGCAACGCTAAGCAACGTAAGCCAT 2553
    : : : : :
856 ..... 856
2554 TCGGCACTCAACGTAATGTCCTCCTAGCCGATTAAGCAGATTCATTT 2603
    : : : : :
856 ..... 856
2604 TGAAGACAGCGCGTTTACGGGACAATCAGCGGCGCAAGATACGGCAT 2653
    : : : : :
856 ..... 856
2654 TACACTTAAAAAGACAGCGAATGAGAGCTGCCGTACGACGCAATTAAGC 2703

```

```

857 .....AsnSerHisTrpHisLeuThrGlyAspSerAsnValAsn 869
      :::::|||||  |||  :::::
2704 AATTTAAACCTTGACAGCGCACATACATCAATCCCGCTATGTGCCA 2753
      :::::|||||  |||  |||  |||  |||  :::::
870 GluLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnAs 884
2754 CGATGCGGCGAGGGCGCAACCGGCAAGTGGCAGATGCGCGCGCGCGC 2803
      :::::|||||  :::::
884 nAspAlaAsnLysValThrThr..... 891
2804 GTTGGCGCGCTGCGCGCGCTCCCTATTCGTTACCGCGCAACTTGC 2853
891 ..... 891
2854 GTAGATCCCGTTTCAACAGCTGACGTAAAGCGCAATTTGAACGTCA 2903
      :::::|||||  |||  |||  |||  |||  :::::
892 .....TyrAsnThrLeuThrValAsn...SerLeuSerGlyAs 903
2904 GGGACATTCCTGCTTATGTGGAAGCTTTCGGCTACCGCAGCACAAT 2953
      :::::|||||  :::::
903 nglySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysV 920
2954 TGAAGCTGGGGAAGTTCGGAAGGCATTAACCTTGGCGTCAACAAT 3003
      :::::|||||  |||  |||  |||  |||  :::::
920 aIValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAsp 936
3004 ACCGGCAGCAAGCTGCAAGCCTCGAACAATTGACGGTAGTAGAAGGAA 3053
      :::::|||||  :::::
937 LysThrGlyLysProThrLys...AsnGluLeuThrLeuPheAspAla.. 951
3054 AGACAAACAACCGCTGTCGAAAACCTTAATTTCAACCTGCAAAAAGAAC 3103
      :::::|||||  :::::
952 ....SerAsnAlaThrAlaArgAsnLeuAsnValSerLeuValGlyAsnT 967
3104 ACCTGCATCCGCGCGCTGACCTTACCAACTCATCCGCAAGACCGGCAG 3153
      |||||  |||||  |||||  |||||  |||||  :::::
967 hrValAspLeuGlyAlaThrLysTyrLysLeuArgAsnValAsnGlyArg 983
3154 TTCCGCTGCTAATATCCG..... 3171
      :::::|||||  |||  |||  |||  |||  :::::
984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000
3171 ..... 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 ..... 3171
1017 rOSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033
3171 ..... 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 ..... 3171
1050 scGlnLysSerLysThrValGluLysAsnGluGlnAspAlaThrGluThr 1067
3171 ..... 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 ..... 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluThr 1100
3172 .....GTCAAGACAACAAGAGCTTCGACAACAACCTGGCAAGC 3208
      :::::|||||  :::::
1100 rGlnThrThrGluThrLysGluThrAlaLysValGluLysGluGluLysA 1117
3209 CAGAAGCCAAAAAACAG..... 3225
      ||:::|||||  :::::
1117 IalysValGluLysGluGluLysAlaLysValGluLysAspGluIleGln 1133
3225 ..... 3225
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150
3225 ..... 3225
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167
3225 ..... 3225
1167 IagInProGlnThrGlnSerThrThrValAlaAlaGluAlaThrSer 1183
3226 .....GCCGAAAAAGACACGC 3242
      :::::|||||  |||||
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200
3243 GCAAAAGCTTGACGCGCTGATTGCG..... 3267
      ||:::  :::::
1200 aGluProValThrProValValSerLysAsnGlnThrGluAsnThrThrA 1217
3268 .....GCCGGCGGAGTCCGCTGCAAAAAGCA 3294
      :::::|||||  |||||
1217 spGlnProThrGluArgGluLysThrThrAlaLysValGluThrGluLysThr 1233
3295 GAAAGCGTTGCGAACCGCGCGCGCAGCAGC..... 3327
      :::::  :::::
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluGlnSe 1250
3328 .....GGGAAAAATGTC..GGCA 3343
      :::::|||||  |||
1250 rGluThrValGlnProGlnAlaValLeuGlnSerLysValAlaProThrV 1267
3344 TTATGACGCGGAGAGCAAGAAAAACGGGTGACGCGATTAAGAC... 3390
      :::::|||||  |||  :::::
1267 aIAsnAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283
3390 ..... 3390
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300
3391 ....ACCGCTTGGCGAAACAGCGGAGCGGAAACCGCGCGCTACC. 3435
      |||||  :::::
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317
3435 ..... 3435
1317 LuThrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3436 .....ACCGCTTCCCGCGCGCGC 3455
      :::::|||||  |||||
1334 AlaAspAsnSerValAlaAsnAsnSerLysSerGluProLysSerAr 1350
3456 CCGCGCCGCGCGGATTTGCCCAACTGCAA..... 3486
      |||||  |||||  :::::
1350 gArg...ArgArgSerIleSerGlnProGlnGluThrSerAlaGluGlu 1366
3486 ..... 3486
1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3487 CCCCAACCGCAGCCCAACCGCAG..... 3510
      :::::|||||  :::::
1383 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399
3511 .....CCGCACTGATCAGCC 3526
      |||||  |||||
1399 lThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuThrSerT 1416
3527 GTTATGCCAATAGCGCTTTCAGTGAAATTTCCGCC..... 3561
      :::::|||||  :::::
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432

```

```

3562 ACGCTCAACAGCGCTTTCCGCTACAGAGCAATTAGACCGGTA...TT 3608
      |||||
1433 AAldeuasnValgllylsalValserGlnHisIleSerGlnLeuLume 1449
      |||||
3609 TGGCCGAGACCGCCGACACCGCTTTGGACAGCGCATCCGGGACACA 3658
      |||||
1449 tAsnAsnGlnGlyGlnIlyrAsnValIlyrValSerAsnThrSerMetAsnG 1466
      |||||
3659 AACACTACCTGTCCAGAGATTTCGCGCCCTACCGCCCAAGAACACCTG 3708
      |||||
1466 IusnIlyrSerSerSerGlnIlyrAlaGlnPheSerSerIlysserThrGln 1482
      |||||
3709 CGCCAAATCGGTATGCAAGAAAACCTCGGACGGG...CGCGTCGGCAT 3755
      |||||
1483 ThrGlnLeuGlnIlyrAspGlnThrIleSerAsnAsnValGlnLeuGlyG 1499
      |||||
3756 CCGCTTTTCCACACCGGACCGAAACACCTTCGACGACGCGATGCGCA 3805
      |||||
1499 yValPheThrIlyrValAlaIyAsnSerAsnAsnPheAspIlyAlaSerSer 1516
      |||||
3806 ACTCGGACGCGCTTCCGACGCGCGCTTTCCGGCAATTCGCGCATCGAC 3855
      |||||
1516 ysaAsnThr...LeuAlaGlnValAsnPheIlySerIlyrIlyrAlaAsp 1531
      |||||
3856 ...AGGTTTACATCGGCAATCAGCGCGCGGCTTTAGCAGCGCGCAG 3902
      |||||
1532 AsnHisIlyrIlyrLeuGlnIleAspLeuGlnIlyrIlyrPheGlnSerAs 1548
      |||||
3903 CCTTTCAGACGCGATCGAGGCAAAATCCGCGCGCGCTGCGATTCAGC 3952
      |||||
1548 nLeuIlyrAsnHisAsnAlaIyAsnPheAlaGlnHisThrAlaGlnPheG 1565
      |||||
3953 GCATTACAGGCGCATACCGCGCGCTTTCCGGGATTCGCGATTCGACGCG 4002
      |||||
1565 IlyLeuThrAlaGlnIlylsalAlaPheAsnLeuGlnAsnPheGlyIleThrPro 1581
      |||||
4003 CACATCGCGCGACGCGCTATTTCGCCAAAAACGCGATTCACCGCTAGA 4052
      |||||
1582 IlyValGlyValAlaIyGlySerIlyLeuSerAsnHisAsnPheAlaLeuAl 1598
      |||||
4053 AAACGTCATATTCGCCACCGCGCGCTTGCATTCACCGCTACCGCGCG 4102
      |||||
1598 aIyAspIlyrIlyrLeuValAsnProIleSerValIlyrThrAlaPheAlaG 1615
      |||||
4103 GCATTACAGGCAATTCATTCATTCAAACCGGCGCAACATTCATTCATCAG 4152
      |||||
1615 InValAspLeuSerIlyrThrIlyrHisLeuGlnGlyGlu...PheSerValThr 1630
      |||||
4153 CCTATTTCAGCGCTGCTCTAT...ACGATGCGCGCTTCGGCGAAAGTCCG 4199
      |||||
1631 ProIleLeuSerAlaIyGlyAspThrAsnGlnIlySerGlyIlySile... 1646
      |||||
4200 AACACGCGCATACCGCGCGCTATTCGCGATTCGCGCAAAACCGCGCA 4249
      |||||
1647 ...AsnValAsnGlnIlyrAspPheAlaIyAsnValGlnIlyAsnGlnGlnG 1662
      |||||
4250 GTGCGGAATGGGCGTAAACGCCGAATCAAAAGTTTCACGCTGTCCCTC 4299
      |||||
1662 InIlyrAsnAlaGlnIlyrLeuIlyrIlyrHisAsnValIlyLeuSerIleu 1678
      |||||
4300 CACGCTGCGCGCGCAAGCGCGCGCATCGCAAGCGCAACACGCGCGG 4349
      |||||
1679 IlyGlyGlnIlyrLeuThrIlyrAsnAlaIySglnAlaIlySglnIlyrThrAlaGln 1695
      |||||
4350 CATCAATTAAGGCTAC 4365
      |||||
1695 uLeuIlyrLeuSerPhe 1700
      |||||
seq_name: p1r2:A81018
seq_documentation_block:
serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogroup

```

```

C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81018
R:RetelIn, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Halt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000, MID:20175755
A:Accession: A81018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <TEI>
A:Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAF42325.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1998

```

```

alignment_scores:
  Quality: 1547.00      Length: 1669
  Ratio: 1.809          Gaps: 50
  Percent Similarity: 51.228      Percent Identity: 27.741

```

alignment_block:

US-09-303-518D-649 x A81018 ..

Align seg 1/1 to: A81018 from: 1 to: 1431

```

55 GCGCGCATCGCTTCGCGCTGCTTACGATTCGCGCTGCTGCG 104
      |||||
17 GlnPheAlaValSerProAlaIySerIleValAlaAsn.... 31
      |||||
105 CATTCCTCCCAAGCCTGCGCGGACACACTTATTCGCGCATCAAC 154
      |||||
32 .....AspValAspTyrG 36
      |||||
155 AATACTATGCGCAGCTTTCGCGAAATAAGCGAGTTTCAGTCGGGCG 204
      |||||
36 InIyPheArGAspPheAlaGlnAsnIlyGlyAlaPheThrValGlyAla 52
      |||||
205 AAGATATTGAGGTTTACAAACAAAGGAGGAGTTCGCGCAATTCAT 254
      |||||
53 SerAsnIleSerIleGlnAspIlySglnGlyIlyleuGlnIyAlaLe 69
      |||||
255 GACAAAAGCCCGATGATGATTTTCTGCTGCTGCT...AACGCGG 301
      |||||
69 uAsnGlyIleProMetProAspPheArGValSerAsnArGlnThrAlaI 86
      |||||
302 TGGCGGATGCGGCGGATTCATATATGTCGCTGCGCATACAGCG 351
      |||||
86 LeAlaThrLeuValHisProGlnIlyrValAsnSerValIlyHisAsnVal 102
      |||||
352 GCGTATACACAGCTGTATTTGGTCGCGAGAGAAATCCGATCAACA 401
      |||||
103 GlnIlyrGlnSerIleGlnPheGlnAsnAspThrGlnAsnProGlnGlnG 119
      |||||
402 TCGTTTACTTATTAATTTGTGAACGAGATATTTAAGCAGGGAGCTA 451
      |||||
119 nAlaIlyrThrIyArGlnValSerArGAsn..... 129
      |||||
452 AAGCGCATGCTTATGCGCGGATTCATATGCGCGCTTTCATTAATTT 501
      |||||
130 ...ProHisPro...AspIlyrAspIlyrHisLeuProAlaGlnAsnIlyLeu 144
      |||||
502 GTCAAGATGCGAGACTGTGAATGACAGT..... 534
      |||||
145 ValThrGlnIlySerProThrAlaLeuSerSerValProLeuGlnGlyAs 161
      |||||
535 ...TATATGATGCGGCGAAATATATCATGATCAAAATATACCTGACC 580
      |||||
161 nGlyGlnProIlylsAlaAsnAlaIyIlyLeuAspPheArGProIlyrP 178
      |||||

```

```
581 GTTCTGATTGGGCGAGCGAGCAATATGCGATCTGATGAGATGAG 630
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 heValArgLeuGlySerGlyThrGlnGlnValArgGlySalAspGly... 193
631 CCCAATTAACCGGAAAGCTTCATATCATATTTGCAAGTCCGATTTCTTGCT 680
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 .....ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205
681 CGTTGGTGCATATACCTTTGCAAAATGGATGAGTGGTGGCGCAAGTCA 730
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 uThrclyGlyThrProLeuYsValLeuGlyPheGlnAsnHisGlyLeuL 222
731 ACTTAGAGTAGTGAATAATTAACATAGCCCAATAGGTTTATTCACAA 780
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 euValGlyGly...SerLeuThrAspLInPro.....LeuAsnThr 234
781 GGAGGCTCATTTGGCGAGTGGCTCCCAATGTTATCTATGATGCCCA 830
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysH 251
831 AAGCAAAAGTGGTAAATTAATGAGGATATGCAACGCGCAACCCCTATA 880
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 sGlnAsnArgTyrValLeuAlaGlyValLeuSerThr.....TyrA 265
881 TAGGAAAAAGCAATGGCTTCAG.....CTGGTTCGTAAGATTGGTTC 924
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 laGlyPheAspAsnPhePheAsnLysTyrLeValThrGlnProGluPhe 281
925 TATGATGAATCTTGGCTGAGATACCCATTCAGTTTCTACAGACCAG 974
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 lLeArgSerThrIleArgGln.....TyrGlnThrTr 292
975 T.....CAAAATGGAAATACCTTTTAAAGCAGATAATA 1009
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 gLeuAspValGlyLeuThrThrAsnGlnLeuIleTyrArgAspAsnGlyA 309
1010 ATGGCAGAGAAATCAATGCCAATGCAACATGATCTCTGCTTAAT 1059
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 snGlyAsnSerThrLeuGlnGlyLeuAsnGlnArgIleThrLeuPro... 324
1060 AGATTAAAAACAGAACCGTTCATTTGTTTATGTTCTTTACCGAGAC 1109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 .....IleAlaAsnProSerLeuAla..... 331
1110 AGCAAGAGAACCTGTTATCATAGCTGCAGGTGGTGTCAACAGTTATTCGAC 1159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 .....P 332
1160 CCAGACTGAATATGAGAAATATTTCTTATTTAGCAAGAGAAA... 1206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347
1207 .....GGCGAATTGATTAATTACACAGAA 1229
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364
1230 CATCAATCAAGTGCTGAGAGATTTATTTCCAGAGATTTTACGGCT 1279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrVal 381
1280 CGCCTGAAATTAACGAAACTTGGCAAGCGCGGCGTTCATATCAGTGA 1329
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 aGlyLysAsnHis...ThrTyrGlnGlyAlaGlyValIleValAlaAsp 396
1330 GACATACCTACTTACTTGAAGTAAAGCGCGTGCACAAAGCGCGCTGTC 1379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 GlyLysArgValPheTyrPheGlnValSerAsnProLysGlyAspArgLeu 413
1380 CAAATTCGCGAAAGCAGCAGCTGCATCAAGCCAAAGGGAACCAACAG 1429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGln 430

1430 GTCATACGCGTGGCGAGCGGTACAGTCAATTTTGATACGAGGAGAC 1479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 lYAspIleSerIleGlyGlnGlyThrValValLeuAlaGlnLysAlaAla 446
1480 GATAAGGCAAAAACAAAGCCTTTAGTGAATAGCGCTTGGTCCAGCGCAG 1529
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyAr 463
1530 GGGTACGGTGCACACTGAATGCCATTAATCAGTTCAACCCCGCAAACTCT 1579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480
1580 ATTTCGCGCTTCCGGCGGAGCTTTGGATTAAACGGGCAATTCGCTTTCG 1629
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 yPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlnLysAsnLeuAla 496
1630 TTCACGCTATTCAAAATACCGATGAAGGGCGCATATGTCACACACA 1679
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 PheThrHisIleArgHisAlaAspGlyGlyAlaGlnIleValAsnHisAs 513
1680 TCAAGACAAAGAAATCCACCGTTCACATTCACGAGCAATTAAGATTGCTA 1729
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 nProAspGlnAlaAlaThrLeuThrLeuThrGlyAsnProValLeuSerP 530
1730 CAACC..... 1734
530 roGluHisValGluTyrValGlnTyrPheGlyAsnArgProGlnGlyAsnAla 546
1734 ..... 1734
547 AlaValTyrGluTyrIleAsnProHisArgAsnArgAlaGlyThrAspTyrP 563
1735 .....GCGCAAT..... 1740
563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580
1741 .....AACAACGCTTGATGACAAA 1761
580 yAsnSerThrSerTyrPheGlnPheIleGlyAsnAsnArgGlnAlaAla 596
1762 AAAGAAATTGCC.....TACAA 1778
597 GluGlnValAlaGlnAlaGlnAsnAlaArgProAspLeuIleThrPheG 613
1779 CGGTGGTGGCGGCAAAAGATACAGC..... 1806
613 yGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAlaProSerTyr 630
1806 ..... 1806
630 eLysThrAsnGlnAlaAlaIleGlyLysThrArgHisIleAlaAsnAla 646
1807 .....AAACGAAGCGGCGCTCAACTTGT 1832
647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuH 663
1833 TTACAGCCCGCCGAGAGAACCGACCGACCTGCTCTTCCGGCGAGCAAA 1882
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 sTyrArgProLysArgThrAspSerThrLeuLeuAsnGlyGlyMetA 680
1883 ATTTAAACGGCAACATCAGCAAAACAAAGGCAAACTGTTTTCAGCGG 1932
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 snLeuAsnGlyGlyValLeuIleGlnGlyLysAsnMetIleValSerGly 696
1933 AGACCAACAGCAGCGCTACCAATCATTTAAACGACCATTTGGTCGCAAAA 1982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
697 ArgProValIleProHisAlaTyr.....AspHisGlnAlaLysAr 709
1983 AGAGGCAATTCCTCGCGGCAAAATCGTGTGGACAACGATCGATCAACC 2032
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
709 gLuePro.....ValLeuGluAsnGluTyrThrAspG 720
2033 GCACATTTAAAGCGGAAACTTCCAATTTAAAGC...GCACAGCGCGGTG 2079
```

```

      :::::::::: ::::: ::::: ::::: :::::
720 lysPheLeuAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736
      :::::::::: ::::: ::::: ::::: :::::
2080 GTTCCCGCAAGTGTGSCAAAGTGAAGGCGAT.....TGCCATTT 2120
      :::::::::: ::::: ::::: ::::: :::::
737 AlagIyArgAsnThrAlaHisLeuAspGlyAspIleThrAlaTyArgPhe 753
      :::::::::: ::::: ::::: ::::: :::::
2121 GAGCAAT...CAGCCCAAGCATTTTGTGTGTCGACCGCATCAAGCC 2167
      :::::::::: ::::: ::::: ::::: :::::
753 uSerGlyIleAspLeuGlyPheThrGlnGlyLeuThrProGln..... 767
      :::::::::: ::::: ::::: ::::: :::::
2168 ACACAACTGTACACGTTGCGACTGACGCGGTGTGACAAATGTGTGCA 2217
      :::::::::: ::::: ::::: ::::: :::::
768 .....CysTyArgSerTyHisSerLeuSerHisSerHisCysThrPro 781
      :::::::::: ::::: ::::: ::::: :::::
2218 AAAACATTCACGACGATAAAGTATGTCCTCATGTGACTAAGACCGCAT 2267
      :::::::::: ::::: ::::: ::::: :::::
782 AsnAlaValIleuLysAlaGlnAsnTyArgAlaIleuProAlaThrGlnVa 798
      :::::::::: ::::: ::::: ::::: :::::
2268 CAGCGCAAGTGCATCTTCCGATCAGCTCATTTAAATCTCACAGGCG 2317
      :::::::::: ::::: ::::: ::::: :::::
798 IArgGlyAspIleThrLeuAsnAspArgSerGlyLeuArgLeu...GlyL 814
      :::::::::: ::::: ::::: ::::: :::::
2318 TTGGCAACCTCAAGCAATCTTGTAGTGCATATGCGATACGCTATACA 2367
      :::::::::: ::::: ::::: ::::: :::::
814 ySalAlaHisLeuTyGlySerIleArgAla..... 823
      :::::::::: ::::: ::::: ::::: :::::
2368 GTCAAGCCAAAGCCACCCAAAGCGACACTTACCTCGTGGCAATGC 2417
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2418 CCAAGCAACATTTAATCAAGCCACATTAAGCGCACATCGGCTTGG 2467
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2468 GCAATGCTCATTTAATCTAAGCAGCAGCCGTAACAAAAGCGCATCTG 2517
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2518 ACGTTTCCGCGACCGCTAAGCGAAGCATTAAGCCATCCGACATCAAGC 2567
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2568 TAATGTCCTCCAGCGGATTAAGCAGATTTCCATTTTGAAGCAGCCGCT 2617
      :::::::::: ::::: ::::: ::::: :::::
823 ..... 823
      :::::::::: ::::: ::::: ::::: :::::
2618 TTACCGGACAAATCAGCGCGGCAAGATACGGCATTAACACTTAAA... 2664
      :::::::::: ::::: ::::: ::::: :::::
824 .....GlyLysAspThrAlaValAlaArgMetGlnAla 833
      :::::::::: ::::: ::::: ::::: :::::
2665 GACAGGGAATGAGCGCTGCCGTCAGGACGGAATTAAGCAATTTAAACCT 2714
      :::::::::: ::::: ::::: ::::: :::::
834 AspSerAsnThrPheThrLeuSerHisSerHisSerHisGlyAlaLeuThrIle 850
      :::::::::: ::::: ::::: ::::: :::::
2715 TGACAGCGCACCATTAACATTTCCGCGCTATCGCAGCATGCGGCGAG 2764
      :::::::::: ::::: ::::: ::::: :::::
850 uAspGlyAlaGlnIleThrLeuAsnProAspPheAlaAsnAsn..... 864
      :::::::::: ::::: ::::: ::::: :::::
2765 GGGCGCAAAACCGGAGTGCAGACAGATGCGCGCGCGCTTCGCGCGCT 2814
      :::::::::: ::::: ::::: ::::: :::::
864 ..... 864
      :::::::::: ::::: ::::: ::::: :::::
2815 TCGCGCGCTTCCCTATTATCCGTACACGCCCACTGCGTGAATCCG 2864
      :::::::::: ::::: ::::: ::::: :::::
865 .....ThrHisAsnAsnArg 869
      :::::::::: ::::: ::::: ::::: :::::
2865 TTTCACAGCGTGCAGTAAACGCAATTTGACGCTCAGGACATTTCC 2914
      :::::::::: ::::: ::::: ::::: :::::
869 gPheAsnThrLeuThrValAsnGlyThrLeuAspGlyPheGlyThrPheA 886
      :::::::::: ::::: ::::: ::::: :::::
2915 GCTTTATGTCGGAATCTTC...GGCTACCGCAGCGCAAAATTTGAAGCG 2961
      :::::::::: ::::: ::::: ::::: :::::

```

[illegible]


```
298 ..GGCGTGGCGCATGTGTGGCGATCAATATATTGTGAGCGTGCACAT 345
      ::::::::::::::::::::|
85 rgrgrhrleuthralleasprroglntyrallavalservallyshis 101
      ::::::::::::::::::::|
346 AACGGCGGTATACACAGCTGATTTGGTGGGAAAGAAACCCCA 395
      |||
102 vallysglyasprleuileretytyrlyshisasnlyshis 118
      ::::::::::::::::::::|
396 T.....CAACATCGTTTACTTAATAATGTGAACGAAATATATA 439
      |
118 pvalserasnaspriusnlytrtyrservalalaglnasnsplyg 135
      ::::::::::::::::::::|
440 AACCGAGACTAAAGCCATCTTATGCGGC.....GAT 474
      ::::::::::::::::::::|
135 lufroasnlyasntrphis...hisglyasnlnlytrleuglnuap 150
      ::::::::::::::::::::|
475 TATCATATGCGCGGTTGCAATAATTTGCACAGATGCAGAACCTGTGA 524
      |||
151 tyasnmetlalarleuasnlysphevalthrleuvalalaproileal 167
      ::::::::::::::::::::|
525 AATGACCACTTATATGATGGC...CGGAATATATCATCAAAATAT 571
      |||
167 aprotthseralaglyglyvalgluthrtyrlyasnplyasnarp 184
      ::::::::::::::::::::|
572 ACCGTGACCGTGTCTATTTGGGCGAGCGCAATATTTGGCATCTGAT 621
      ::::::::::::::::::::|
184 heserleuphevalarvalgllyalcllythrlnphe..... 196
      ::::::::::::::::::::|
622 GAAGATGAGCCCAATACCGGAAAGTTTCATATATTCATCAAGTGGTA 671
      |||
197 .....gluthrasnserarglytyrasmethrclueuaserarvalty 211
      ::::::::::::::::::::|
672 TTCTTGCTCGTGGTGGCAATACCTTT.....GCACAA 706
      |
211 rargtyralallealcllythrprotyrlnasprvalasnvalthrsera 228
      ::::::::::::::::::::|
707 ATGATCATAGTGTGGCAGCTGCACTAGGTAGTGAATAATTAACAT 756
      |||
228 snleuasnlnlyleuileglypheglyasprasnserlyshis 244
      ::::::::::::::::::::|
757 AGCCCA.....TATGTTTTTTACCAACAGG 782
      |||
245 Serprogluylsleuylslnvalleuaserlnasnaleuthrasnty 261
      ::::::::::::::::::::|
783 AGGCTCATTTGGCGACAGTGGCTACCAATTTATCATGATGCCCAAA 832
      ::::::::::::::::::::|
261 ralavalleuylasprserglyserproleuphealalyrasplysgln 278
      ::::::::::::::::::::|
833 AGCAAAAGTGGTAAATTAATGAGGTATGCAACGGGCAACCCCTATATA 882
      ::::::::::::::::::::|
278 lulyasrtyrvalrphelenuglyalalyraspyrtyrpalaglytyr... 293
      ::::::::::::::::::::|
883 GGAANAAGCAATGCGTCCAGCTGGTTCGTAACATTTG...TTCTATGA 929
      ::::::::::::::::::::|
294 ...glulyasnsertrpdl.....glutrasnlietyrly 305
      ::::::::::::::::::::|
930 TGAATCTTTGCTGGAGATACCCATTCAGTATTTACAGACCCGTCAAA 979
      ::::::::::::::::::::|
305 slusleuphealasp.....glulleysglna 315
      ::::::::::::::::::::|
980 ATGGGAATACTTTTAAAGCATTAATAATGCGACAGAAATTCAT 1029
      ::::::::::::::::::::|
315 rg.....Aspasnalaglythrleelsglytyr 324
      ::::::::::::::::::::|
1030 GCCAAACATGAACACATTTCTGCTAATAGATTAACAGCAAGCGT 1079
      ::::::::::::::::::::|
325 gluglyshisrtyrlyshis.....Thrglythr... 334
      ::::::::::::::::::::|
1080 TCAATTTGTTATGTTCTTTATCCGACAGACAGAGAACCTGTTTATC 1129
      |||
335 .....AsnserhisileglyserthrlyalavalrleuvalaglyA 348
      ::::::::::::::::::::|
1130 ATGTGACAGGTGTGTCAACAGTTATGCAACCCAGCTGAATATAGGAGA 1179
      ::::::::::::::::::::|

      ::::::::::::::::::::|
348 snlgluaraglyala.....Asnasnlygln 356
      ::::::::::::::::::::|
1180 AATATTTCTTTATTTGACGAGAGAAAGCGGANTGTACTTACACAA 1229
      ::::::::::::::::::::|
357 Asnvalthrpe.....Gluasnasnlythrleuvalleuasprius 371
      ::::::::::::::::::::|
1230 CATCAATCAAGTCTGAGATATATTTCCAGGATTTTACGGTCT 1279
      |||
371 nllasnlncllyalaglyglyleuphepnepneglyasprtyrthrval 388
      ::::::::::::::::::::|
1280 CGCCTGAATAATACGAA...ACTTGCAAGCGCGCGCTTCATATACGT 1326
      ::::::::::::::::::::|
388 ysglyalasnasnlylthrthrleuaglyalaglyleaspralala 404
      ::::::::::::::::::::|
1327 GAAGCACTACCGTTACTTGAAGTAAAGCGCGTGCAGAACGCCCT 1376
      ::::::::::::::::::::|
405 asprglylysvalvaltyrprlvallyasnproasnlyaspralagly 421
      ::::::::::::::::::::|
1377 GTCCAAATCGGCAAGGACGCTGCACGTTCAAGCCAAAGGGAACC 1426
      ::::::::::::::::::::|
421 ualalyshileglylysglythrleuaglyleasnlythrlyalasn 438
      ::::::::::::::::::::|
1427 AAGCTCGATCAGCGTGGCGACGCTACAGTCATTTTGATCAGCAGCA 1476
      |||
438 lnglyshleuylsvalglyasprglythrvalilleleuasnlnlalnla 454
      ::::::::::::::::::::|
1477 GACGATAAAGCAAAACCAAGCCTTTAGTGAATCGGCTGTGACCGG 1526
      |||
455 aspralasnplyslyvalgllnalrpheserlnvalgllylevalsercl 471
      ::::::::::::::::::::|
1527 CAGGGTACGCTGCAACTGAATCCGATATCATCTTCAACCCGCAAC 1576
      |||
471 yarglythrleuvalleuasnserproasprlnleasnproasnaln 488
      ::::::::::::::::::::|
1577 TCTATTTGCGCTTTGGCGGCGACGCTTGGATTTAAAGGCGATCGCTT 1626
      |||
488 eutyrrheglyphearaglyglyargleuaspalasnlnlyasnsprieu 504
      ::::::::::::::::::::|
1627 TCGTTCACCGTATTCAAATACCGATGCAAGGGGAGATGATTTGCAACA 1676
      ::::::::::::::::::::|
505 thrpheglnhislearghsnvalasprluclyalalargllevalasnhi 521
      ::::::::::::::::::::|
1677 CAATCAAGCAAAAGANTCCAGCTTACCTTACAGC..... 1713
      |||
521 saenthrassparvalaserthrleuthrleuglyysseulleier 538
      ::::::::::::::::::::|
1713 ..... 1713
      ::::::::::::::::::::|
538 hrAlaProglinsnleuSerValTyrclullearghasnsplyraspser 554
      ::::::::::::::::::::|
1714 .....ANTPA 1718
      ::::::::::::::::::::|
555 AspAspTyTyrGlyTyrTyrSerTyTyrGlyspProileProglInGly 571
      ::::::::::::::::::::|
1719 AGATATT..... 1725
      ::::::::::::::::::::|
571 saspleutyrrtyrlyasnlytyrtyrlyalaleuylsaserglylys 588
      ::::::::::::::::::::|
1726 .....GTCACACCGGCANTAAC..... 1743
      ::::::::::::::::::::|
588 erylalasnAlaPromelProgluasnnglylnthrcllunasnnaasprtr 604
      ::::::::::::::::::::|
1744 .....AACAG 1748
      ::::::::::::::::::::|
605 lleuemetglyserthrnglnluclyualalalylysasnAlaMetasnhi 621
      ::::::::::::::::::::|
1749 CTTCGATAGCAAAAGAAATTCGCTACACAGGTGTTGTCGAGAAAG 1798
      ::::::::::::::::::::|
621 slvsasnasnlnarglyleserglypheserglyrpheseglygluclyua 638
      ::::::::::::::::::::|
1799 ATACGACCAAAACGACGCGGCTCAACCTTGTATACAGCCGCCGCGCA 1848
      ::::::::::::::::::::|
```

```

638 snGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPhesnGlyLysSer 654
1849 GAAGACGGCACCCCTGCTGCTTCCGGGGAGCAAAATTAAACGGCAACAT 1898
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
655 AlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyLys11 671
1899 CACGCAACAAACGAACTGTTTTCAGCGGCAACACACACGGCAGC 1948
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
671 eSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisA 688
1949 CTTACAAATCATTTAAAGACCATTTGGTCCGCAAAAGAGGGC.....ATT 1992
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
688 laArgAspPheValAsnLysSerSerAlaGlnLysAspAlaHisPheSer 704
1993 CTTCCGGGGGAAATCGTGTGGACAACAGCTGATCAACCGCACATTTAA 2042
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
705 LysAsnAsnGlyValAlaPheGlnAspAspTrpIleAsnArgThrPheYl 721
2043 AGCGGAAACCTTCCAAATTAAAGCGGAGACGGCGCTTCC...CGCA 2089
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
721 sAlaThrGlnIleThrValAsnGlnSerAlaSerPheSerSerGlyArgA 738
2090 ATGTTCGCAAAAGTGAAGGCGATTGGCATTTGACATCACCGCCACGA 2139
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
738 snValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysVal 754
2140 GTTTTGGTGTGGACCGGCATCAAAACCAACATCTGTACACGTTCCGA 2189
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
755 AsnLeuGlyTyr.....LysAsnGlyAspGlyValAlaArgSerAs 769
2190 CTGGACGGGCTGACAAATTTGTGGAAAAACCATTAACCCAGTAAG 2239
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
769 pArgThrGlyTyrValThrCysAsnThrGlyAsnLeuSer...AspLysA 785
2240 TGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGATCTGTGCC 2289
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
785 laLeuAsnSerPheGlyAlaThrLysIleAsnGlyAsnValAsnLeuAsn 801
2290 GATCAGCGCATTTAAATCTCAGAGGCTTCCACACTCAACGGCAATCT 2339
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
802 GlnAsnAlaAlaLeuValLeu...GlyLysAlaAlaLeuTrpGlyGln11 817
2340 TAGTCAAAATGGCGATACAGCTTATACGTCAGCACACAGCC..... 2382
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
817 eGlnGlyGlnLysAsnSerArgValSerLeuAsnGlnHisSerLysTrpH 834
2383 .....ACCCAAAAGCGCAACCTTACGCTC..... 2406
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
834 isLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHis 850
2407 .....GTGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
851 IleHisLeuAsnAsnAlaSerAspAlaGlnSerAlaAsnGlnLysHisTh 867
2442 ATTTAAACGGCAACATCGGCTTGGGCAATGCTTCATTTAACTAAGCG 2491
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
867 rLeuLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisIstLysLeuT 883
2492 ACCACGCCGCTACAAAC...GGCAGTCTGACGCT..... 2523
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
883 hrHisLeuAlaGlnLeuAlaGlnLysAspLysValLeuValLysGlnSerAla 899
2524 TCCGGCAACGCTAAGCAACGTAAGCCATTCCGACACTCAACGGTAAT.. 2571
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
900 SerGlyHisIstLysGlnLeuHisValGlnAspLysThrGlyLysProAsnGln 916
2571 ..... 2571
916 nGlnGlyLeuAsnLeuPheAspAlaSerValAlaArgAspArgSerHisL 933
2572 .....GTCTCCCTAAGCGATAAGCGATATTCATTTTGAAGCAGCGCG 2616
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
933 euSerValSerLeuAlaAsnHisVal...AspLeuGlyAlaLeuAsnArg 948
2617 TTTCACGGACAAATCAACGGCGGCAAGATACGGCTTACACTTAAGA 2666
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
949 TyrThrIleLysThrGlnAsnGly...IleThrArgLeuTyrAsnProTy 964
2667 CACGGAA.....TGACGCTGCCGCTGACGACGG 2695
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
964 rAlaGlnAsnArgArgValLysProAlaProSerProAlaThrAsnT 981
2696 AATTAGCAATTTAAACCTTGACACGCCACCATTA..... 2733
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
981 hrAlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 997
2734 .....CTCAATTCGCGCTA 2747
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
998 IleValValAlaProProSerProGlnAlaAsnGlnAlaGlnLysAla 1014
2748 TCCGCACAGATGCCGCGGCGGCGCAAAAC..... 2775
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1014 sArgGlnGlnAlaLysAlaGlnGlnValLysArgGlnGlnAlaGlnLag 1031
2775 ..... 2775
1031 LuArgLysSerAlaGlnLeuAlaLysGlnLysAlaGlnLysAlaGlnLag 1047
2776 .....GGCAGTGGCAC 2786
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1048 AlaArgGlnLeuAlaThrArgGlnLysAlaGlnGlnLysArgSerAl 1064
2787 AGATGGCGCGCGCCGCTGCGCGCGCTGCGCGCTTCCCTATTATCCG 2836
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1064 aGlnLeuAlaArgArgHisGlnLysGlnArgGlnAlaLagLysLeuSerA 1081
2837 TTACACCGCAACTTCGCTAGAACCTGTTCAACACGCTGACGCGTAAC 2886
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1081 laLysGlnLysValGlnLysAlaGlnArgGlnAlaLysAlaValArg 1097
2887 GCGCAATTGACGGTCAGGAGCAACATTCGCTTATGTCGCAACTCTT..CG 2935
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1098 ArgLysAlaGlnAlaGlnLysArgGlnAlaLysArgGlnAlaLagLysLeuAlaTr 1114
2936 GCTACCGGACGACAAATTGAAGCT..... 2960
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1114 gArgHisGlnLysGlnArgGlnAlaLysAlaGlnLeuSerAlaLysGlnArgY 1131
2961 ..GGCGAAAGTTCGGAAGCATTACACCTTGGCGGCTCAACAATACCG 3008
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1131 aGlyGlnGlnLysArgGlnThrAlaGlnSerGlnProGlnArgArg 1147
3009 CACGCAACCTGCAAGCCTCGACACAAATTCAGCGTATGTGAAG..... 3050
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1148 LysArgAlaAlaAlaProGlnAspTyrMetAlaAlaSerGlnAspArgTr 1164
3050 ..... 3050
1164 OlyArgArgGlyHisArgSerValGlnGlnAsnAsnValGlnIleAlaG 1181
3051 .....AAAAGACACAA..... 3062
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1181 lAlaGlnAlaGlnLeuAlaArgArgGlnGlnGlnLysAlaAla 1197
3062 ..... 3062
1198 GlnLeuLeuAlaLysGlnArgAlaGlnLysAlaGlnLysArgGlnAlaLag 1214
3063 ..ACCGCTGTCGCAAAACCTTAATTTCACCCCTCAAGCAACGCGCAT 3111
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1214 uAlaAlaLysArgLysAlaGlnLysAlaGlnLysAlaLysArg..GlnAlaAla 1230
3112 GCGCGCGCTGGCGGTTACCAACTCATTCGCAAGACGGCGAGTTCCGCT 3161
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1231 GlnLeuAlaHisArgGlnLysAlaGlnLysArgLysAlaAlaGlnLeuSerAl 1247

```

```

1542 aglnlyrgharghpheserSerlysrghrlnthnglnlleglylaa 1559
      :::::::::::::: :::::::::::::: ::::::::::::::
3725 AGAANAACCTGGGACGGGG...CGCGTGGCATCTGTTTCGCACAC 3771
      :::::::::::::: :::::::::::::: ::::::::::::::
1559 spargserleuserSerGuaasmetGlnlleglylvalleuthrlyrSer 1575
      :::::::::::::: :::::::::::::: ::::::::::::::
3772 CGGACCGGAAACACCTTCGACGACGATCGGCACTCGGACGGCTTGC 3821
      :::::::::::::: :::::::::::::: ::::::::::::::
1576 AspserGlnlsthphespglnAlaserglyLysasnThr...pheya 1591
      :::::::::::::: :::::::::::::: ::::::::::::::
3822 CCACGGGCGCGTTTTCGGGCAATACGGCATC...GACAGCTTACATCG 3868
      :::::::::::::: :::::::::::::: ::::::::::::::
1591 lGlnAlaasnleuthrGlylystrlyrlyLeuaasnsPalatrpyrmetA 1608
      :::::::::::::: :::::::::::::: ::::::::::::::
3869 GCATCAGCGCGGGCGCGGTTTACGACGGCAGCCCTTCAGACGGCATC 3918
      :::::::::::::: :::::::::::::: ::::::::::::::
1608 lnelysprrleglyAlaaglySerleuarSerargleuglnthngln 1624
      :::::::::::::: :::::::::::::: ::::::::::::::
3919 GGAAGCGAAATTCGCGCCGCGCTGCATTCAGGCAATTCAGGACAGATA 3968
      :::::::::::::: :::::::::::::: ::::::::::::::
1625 LysAlaasnphesasnArgThrSerlleGlnthnglyLeuthrleuglyas 1641
      :::::::::::::: :::::::::::::: ::::::::::::::
3969 CCGCGCGCGTTTCGGCGGATTCGGCATCGAAGCAGCATCGCGCAACGC 4018
      :::::::::::::: :::::::::::::: ::::::::::::::
1641 nThrLeuylsleasnGlnPhegluilevalProserAlaglyLeargt 1658
      :::::::::::::: :::::::::::::: ::::::::::::::
4019 GCATTTTCGTCAGAAAGCGGATTCGCGTACGAAACGTCAATATTCGCC 4068
      :::::::::::::: :::::::::::::: ::::::::::::::
1658 yrserArgleuserSerAlasprrlyrlyleuglyAsnAspSerAllys 1674
      :::::::::::::: :::::::::::::: ::::::::::::::
4069 ACCCGCGCGCTTCGATTCACACCGCTACCGCGCGGCAATTAAGCAGATTA 4118
      :::::::::::::: :::::::::::::: ::::::::::::::
1675 ValsermetSerValylsThrleuthrAlaelyleuasnspPhealaty 1691
      :::::::::::::: :::::::::::::: ::::::::::::::
4119 TTCATTTCAACCGGCGCAACACATTTCCATTCAGCGCTTATTTGAGCCTGT 4168
      :::::::::::::: :::::::::::::: ::::::::::::::
1691 rArgPheyls...ValglyAsnleuthrValylsPheoleuSerAlaa 1707
      :::::::::::::: :::::::::::::: ::::::::::::::
4169 CCTAT...ACCGATCCCGCTTCGGGAAAGGCCGACGCGGCAATATTCAC 4215
      :::::::::::::: :::::::::::::: ::::::::::::::
1707 lathrPheAlaasnlyrGlylysglyGlyAlaasnValGlyGlyAsnSer 1723
      :::::::::::::: :::::::::::::: ::::::::::::::
4216 GCCGATATTCGCTCAGATTTTCGGCAAAACCGCGCATGCGGATGGCGGT 4265
      :::::::::::::: :::::::::::::: ::::::::::::::
1724 PheValylrlyrAlaaspAsnGlnGlnGlnThrSerAla.....GlyAl 1738
      :::::::::::::: :::::::::::::: ::::::::::::::
4266 AAGCGCGCAATCAAGGTTTCAGCTTCCTCCCTCCACCTCGCGCGCGCA 4315
      :::::::::::::: :::::::::::::: ::::::::::::::
1738 alaLeuLeuthrArgasnValThrleuasnValasnGlySerllethrL 1755
      :::::::::::::: :::::::::::::: ::::::::::::::
4316 AAGCGCGCACTGGAAGCGCAACACAGCGCGGATCAATTAAGGCTAC 4365
      :::::::::::::: :::::::::::::: ::::::::::::::
1755 ysglyLysGlnleuGlnlysglnlysserglyGlnlleyslleGlnlle 1771
      :::::::::::::: :::::::::::::: ::::::::::::::
4366 CGCTGG 4371
      :::::::::::::: :::::::::::::: ::::::::::::::
1772 Argphe 1773
      :::::::::::::: :::::::::::::: ::::::::::::::
seq_name: pLR2:C81169

seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - Neisseria meningitidis
C|Species: Neisseria meningitidis
A|Variety: strain MC58 serogroup B, strain 81139
C|Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C|Accession: C81169; S61326
R|Retelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rt, H.; Qin, H.; Vamthepan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizzo, M.
Science 287, 1809-1815, 2000
A|Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

```

A:Reference number: A81000; MUID:20175755
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <TEP>
A:Cross-references: GB:AE002424; GB:AE002098; NID:97225923; PIDN:AP41117.1; PID:9722592
A:Experimental source: serogroup B, strain MC58
R:Lehmlolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria*
A:Reference number: S61314; MUID:95502961
A:Accession: S61326
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82477; NID:9732856; PIDN:CA57860.1; PID:9732857
C:Genetics:
A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 1491.50 Length: 2000
Ratio: 1.637 Gaps: 56
Percent Similarity: 45.550 Percent Identity: 24.550

alignment_block:

US-09-303-518d-649 x C81169 ..

Align seg 1/1 to: C81169 from: 1 to: 1815

```
49 AAGAGCGGCGCATCGCTTCGCGCTGCTACTTACGCAATAGCGCTGTC 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATTTCTCCCAAGCCTGGCGGACACACTATTTCGGCATCA 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 atyralaleuthrprotyrserglualalaleuvalalargaspvala 35
149 ACTACCAATACTATCGCGACTTGGCCGAATAAAGGCAAGTTGCACTG 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 spyrctlinlepheargasphealagluasnlysglyllysphephea 51
199 GGGGCGAAGATATGAGGTGTACACAAAGGAGGAGGTGGTGGCGCAA 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 glualalathaspheuservallysasnlysglnlyglnasnilegly 68
249 ATCAATGACAAAGCCCGATGATGATTTCTGTGTGTCGCGGTAAAC 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 malaleuserasnvalprometileaspheserValalalaspyalasn 85
298 ..GGGTGGCGCATTTGGTGGCGCATCAATATATTTGAGCGTGCACAT 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 rgarthrleuthrvalilleasproglntyrvalalaserVallysnis 101
346 AAGCGGCGCTATTAACAAGCTTGTGTTGGCGGAGGAGGAGAAATCCCG 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 vallysglyaspgluileserTyrtyrcllyhisasnlysnleuas 118
396 T.....CAACATGTTTACTTAAATTTGTAAGGAGAAATATATATA 439
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 pvalserasnaspelunsnlntyrargserValalaglnasnasprryc 135
440 AAGGAGGACTAAAGGCACTCTATAGCGGC.....GAT 474
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 luproasnlysnasntprhis...Hiselysnlnlntylargleuclunsp 150
475 TATCATATGCCGCGTTCATTAATTTGTCAGATGACAGAACCTGTGTA 524
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 tyrsnmetalaargleuasnllysphevalthrclvalalaproileal 167
525 AATGACAGTATATATGATGGG...CGGAATATATGATCAAAATATAT 571
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
167 aProthSerAlaGlyGlyValGluThrTyrLysAspLysAsnArgp 184
572 ACCCTGACCGGTTCGATTTGGGCGACGACGCAATATTTGGCGATCTGAT 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 heserGluphevalArgvalGlyAlaGlyThrclInphe..... 196
622 GAAGATAGCCCAATACCGCAAGAGTTATCATATGATGCAAGTGGCGTA 671
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTy 211
672 TTCTGGCTGCTGGTGGCATACCTTT.....GCACAA 706
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 rargtyralalaleaglythrprotyrglnaspvalasnvalthrsera 228
707 ATGATCAGGTGGTGGCGACGACGCACTTAGTAGTGAAGAAATTAACAT 756
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228 snleuasnglnlntylleuileglyphecllyaspasnseryshishs 244
757 AGCCCA.....TATGTTTTTACCAACAG 782
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
245 SerProGluLysLeuLysgluValleuSerGlnAsnAlaleuthrAsnTy 261
783 AGCTCATTTGGCGACGATGGCGCACCAATGTTATTCATGATGCCAA 832
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 ralaValleuGlyAspsercllySerProleuPhealatyrsplysgln 278
833 AGCAAAAGTGTATTAATTAATGGGGATTTGCAAACGGCACCCCATATA 882
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
278 lulsarTyrpvalPheleuGlyAlaTyAspTyrThrAlaGlyTyr... 293
883 GGAAGAAAGATAGCTTCACAGCTGTTGTGAAGATTG...TTCTATGA 929
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
294 ...GlnLysAsnSerTrpeln.....GluTrpAsnIleTyrLy 305
930 TGAATCTTTGCTGAGATACCATTCAGTATTCACACAGACAGTAA 979
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305 stysgluPhealasp.....GluIlelysglna 315
980 ATGGGAATACCTCTTTAACGACATTAATGCGACAGGAAGAAATCAAT 1029
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 r9.....AspasnAlaGlyThrIlelysglyAsn 324
1030 GCCAAACATGAACACATTCCTGCTGATATGANTTAACACACACACCT 1079
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
325 GlyGlnHisHisThrLysThr.....ThrGlyThr... 334
1080 TCAATTTGTTAATGTTCTTATCCGACACAGCAAGAACCTGTATTC 1129
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
335 .....AsnSerHisIleGlySerThr..... 341
1130 ATGCTGAGGTGGTGTCAACAGTTATGACCCACAGCTGAATATGAGAA 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 ..AlaValalargleuAlasasnlnlntargasp...Alaasnanglyln 356
1180 AATATTTCTTATTTGACGAAGAAAGCGAATGATGATCTTCCAGCAA 1229
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 AsnvalThrPhe.....GluAsnanglyThrleuValleuasplnns 371
1230 CATCATATAGGTGCTGAGAGATTAATTTTCAAGAGATTTAGCGTCT 1279
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 nileasnlnlntylalaglyleuPhePhePheLyAspTyrThrVal 388
1280 CGCCTGAATAATACGAA...ACTTGGCAAGCGCGCGCTTCAATCACT 1326
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 ysglyIleAsnAsnaspIleThrThrleuGlyAlaGlyIleAspValAla 404
1327 GAACACATGACCTTACTTGAAGAAAGTAAACGGCTGCAAGACGCGCT 1376
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 AspelyLysLysValalThrGlnValLysAsnProAsncllyAspArgle 421
1377 GTCCAAATATGCGCAAGGCGCTGACAGTTCACAGCAAGGCAAAACC 1426
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 uAlaLysIleGlyLysGlyThrleuGlnleuasnglyThrGlyValasn 438
```



```

171 uasplysheValThrGluValAlaProLeuLeuAlaSerThrAlaAsn 188
172 |||||..... |||||.....
542 ATGGCGG...AAATATGATCAAAATATATACCTGACCGTGTCT 588
188 snAsnLysGlyIuTyraSnAsnSerAspLysTrpAlaPheValArg 204
589 ATTGGGCGAGCAGCGCAATAT..... 609
205 LeuGlySerGlySerGlnPheIleTyrlsGlySerArgTyGlnLe 221
610 .TGCGCATCTGATGAAGATGAGCCCAATATACCGCAAGTTCATATCAT. 657
221 uileuThrGluLysAspLysGlnIleLysLeuLeuArgAsnTrpAspV 238
658 .....ATGCAAGTGGCATCTGCTGCTGCT 684
238 aLgLyLysAsnLeuGluLeuValGlyAsnAlaIleTyrlsGlyIle 254
685 GGTGGCAATACCTTTCACAAATGATGAGTGGTGCACAGTCAACT 734
255 AlaGlyThrProTyrlsValAsnHisGluAsnAsnGlyLeuIleGlyPh 271
735 AGTAGTGAATAATTAACATAGC...CCATATGCTTTTACCACAG 781
271 eGlyAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnA 288
782 GA.....GGCTCATTTGGCGACAGTGCCTCACCAATGTT 816
288 sPrProLeuThrAsnTyrlaValLeuGlyAspSerGlySerProLeuPhe 304
817 ATCTTGATGCCCCAAAGCAAAAGCTGTAATTAATGGGTATTCGAAC 866
305 ValTyraPheArgGluLysGlyLysTrpLeuPheLeuGlySerTyraPhe 321
867 GGGCAACCCCTATATAGAAAAGCAATGCTTCACGCTGCTGCTGAAG 916
321 eTrpAlaGlyTyraLeuLysSerTrpGlnGluTrpAsnIleTyrlsH 338
917 ATTGGTCTATGATGAATCTTTCGTGAGATCCCATCGATTCATTCAC 966
338 IsGluPheAlaGluLysIleTyrls.....GlnGlnTyrlsAlaGlySer 352
967 GAACCACTCAAAATGGAAATACTCTTTTACGACGATTAATATGCGCAC 1016
353 LeuThrGlySerAsnThrGlnTyrlsTrp.....GlnAlaThr 365
1017 AGGAAAAATCAATGCCAAACATGACAACTCTGCGCTAATAGATTAA 1066
365 rGly.....S 367
1067 AAACACGAAACCGT.....CAATTTGTTAATGTTTCTTATATCC 1104
367 eThrIleThrIleThrGlyGlyLeuProLeuSerValAlaPheLeuThr 383
1105 GAGACACGACGAACTGTATTATCATGCTGAGAGTGTCTCAACAGTAA 1154
384 AspLysLysAspLysPro..... 389
1155 TCGACCCAGACTCAATTAATGAGAAAATATTCTTATTATGACGAAGAA 1204
390 .....AsnHisGlyLysSerIleThrLeu.....LysGlyLys 400
1205 AAGCGCAATGATACCTTACACGACATCATCAAGTGTGAGAGATTA 1254
400 eGlyThrIleThrLeuAsnAsnHisIleAspGlnIleAlaGlyLeuLeu 416
1255 TATTTCAAGGAGATTTTAACGTC...TGCGCTGAATAATACGAACCTTG 1301
417 PhePheLeuLysAspTyrlsValLysGlyThrSerAspSerThrThrTr 433
1302 GCAAGGCGCGGCTTCATATGATGAAAGACGATACCGTTACTTGAAG 1351
|||||..... |||||.....
433 pLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysV 450
1352 TAAACGGCGTGGCAAGAACCGCTGTCCAAAATGGCAAGACACCTG 1401
450 aLHisAsnProLysTyraPheArgLeuAlaLysIleGlyLysGlyThrLeu 466
1402 CACGTTCAACCCAAAGGGGAAAACCAAGCTGCATCAGCTGAGGGGAGG 1451
467 ValValGlnIuGlyLysAsnGlnGlyLeuLeuLysValGlyAspG 483
1452 TACAGTCATTTTGGATCAGCAGCAGACAGATTAAGCAAAAACAAAGCT 1501
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaP 500
1502 TTAGTAAATCGGCTGTGTCAGCGAGGCTGACGAGCTGCACATGATCC 1551
500 hSerGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAsp 516
1552 GATATCACTTCAACCCGCAAACTCTATTGCGCTTTCGCGCGAGC 1601
517 AspLysGlnValAspProAsnSerIleTyrlsPheGlyPheArgGlyArg 533
1602 TTTGATTTAAACGGGCAATTCGCTTTCGTCCACCGTATTCAAAATACG 1651
533 gLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleA 550
1652 ATGAAGGGCGATGATTTGTCACACCAATCAAGCAAAAGATCCAGCT 1701
550 sPAspGlyAlaArgValAlaAsnHisAsnMetThrAsnThrSerAsnIle 566
1702 ACCATTACAGCG.....ATAAAGATTTGCTACACCGG 1736
567 ThrIleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTy 583
1737 CAAT..... 1740
583 rAsnIleGluAlaGlnAspAspHisProLeuArgIleArgSerIleP 600
1740 ..... 1740
600 rOTyrlsArgGlnLeuTyrlsPheAsnGlnAspAsnArgSerTyrlsThrLeu 616
1740 ..... 1740
617 LysLysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyLysE 633
1740 ..... 1740
633 rAsnGluAsnTrpLeuTyrlsMetGlyArgThrSerAspGluAlaLysArg 650
1741 .....AAACACAGCTGTGATGCAAAAAGAAATTCGCTACAAACGCTTG 1785
650 snValMetAsnHisIleAsnAsnGlnTyrlsMetAsnGlyPheAsnGlyTyrls 666
1786 TTTGGCGAAGAAATACGACCAAAACG...AACGGCGGCTCAACCTTGT 1832
667 PheGlyGluGlnIuThrLysAlaThrGlnAsnGlyLysLeuAsnValThr 683
1833 TTAACACCCCGCGCGAGAGACCGCAACCTGCTGCTTTCGCGGAGACA 1882
683 rPheAsnGlyLysSerAspLeuAsnArgPheLeuThrGlyGlyThrA 700
1883 ATTAAACGGCAACATCAGCAAAACGCAAACTGTTTTCAGCGCG 1932
700 snLeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerLys 716
1933 AGACCAACCGCGAGCGCTACATCATTTAAACGACCAATGGTGGCAAA 1982
717 ArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLys 733
1983 AGAGGCG.....ATTCTCGCGGGGAAATCGTGGGACAAACGCTGGA 2026
733 sAspProHisPheThrGluAsnAsnGluValValValGluAspAspTrp 750

```


1754 ATAGCAAAAAGAAATTCCTACACGGTTG..... 1785
538LysValAlaLeuasnGlyTrpSerGlyLysGly 550
1785 1785
551 ThrAlaGlyAsnLeuTyrLysTyrAsnAsnProTyrThrAsnThrAsn 567
1786TTTGGCGAAGATAGTACGAC..... 1806
567 pTyrPheLeuLeuLysGlnSerThrTyrGlyTyrPheProThrAspGlns 584
1806 1806
584 erSerAsnAlaThrTrpGluPheValGlyHisSerGlnGlyAspAlaGln 600
1807AA 1808
601 LysLeuValAlaAspArgPheAsnThrAlaGlyTyrLeuPheHisGlyGly 617
1809 AACGACGGGGGCGCTCAACCTGTTCACGAG...CCCGCGCGAAGAGAC 1855
617 nLeuLysGlyAsnLeuAsnValAspAsnArgLeuProGlnGlyValThrG 634
1856 GCAACCTGCTGCTTCGCCGCGAACAATTTAAACGGCAATCAGCA 1905
634 LysAlaLeuValMetAspGlyAlaAlaAspLysSerGlyThrPheThrGln 650
1906 ACAACGGCAACTGTTTTCAGCGCGACAGCAACCGCAGCGCTACAA 1955
651 GluAsnGlyArgLeuThrLeuGlnGlyHisProValIleHisAlaTyrAs 667
1956 TCATTTAAAC.....GACCAT..... 1971
667 nThrGlnSerValAlaAspLysLeuAlaAlaSerGlyAspHisSerVal 684
1972TGTCGCAAAAAGAGCGCATTCCTCGCGGCA 2004
684 euthrGlnProThrSerPheSerGlnGlu..... 693
2005 ATCGTGTGGACACGACTGATCAACCGCACATTTAAAGCGAATACT 2054
694AspTrpGlnAsnArgSerPheThrPheAspArgLeu 705
2055 CCAATTTAAGCGGACAGCGCGGTTCCTCCGCAATGTTGCCAAAGTGA 2104
705 uSerLeuLysAsnThrAspPheGlyArgAsnAlaThr..... 719
2105 AAGCGATGGCATTTGACGATTCAGCGCAAGCA...GTTTGTGCTC 2151
720LeuAsnThrIleGlnAlaAspAsnSerSerVal 731
2152 GCACCGCATCAAGCCACACATCTGTACAGTTTCGACTGAGCGGTCT 2201
732 ThrLeuGlyAspSerArgValPheIleAspLysAsnAspGlyGlnGlyTh 748
2202 GACAATTTGTGTGCAAAA.....ACCATTAACGACGATTAAGTGA 2242
748 rAlaPheThrLeuGlnGlyGlyThrSerValAlaThrLysAsp..... 762
2243 TTGCTTCACTAGACGACGATTCAGCGCATTCGATCTTCGGCAT 2292
763AlaAspLysSerValPheAsnGlyThrValAsnLeuAspAsn 776
2293 CACGCTCATTTAAATCTCACAGGGCTTGCACACTCAACGGCAATCTTAG 2342
777 GlnSerValLeuAsnIleAsnAspIle.....PheAsnGlyGlyIleGly 791
2343 TGCAAATGGCATACAGTTATACAGTCACCAACGCCACCAACCAAG 2392
791 nAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804
2393 GCAACCTTAGCTCGTGGGCAATGCCAAGCAACATTTAATCAACGACA 2442
805ValLeuGlyAsnSerThrLeuThrSerThrAlaLeuAsn 817
2443 TTAACGGCAACATCGGCTTCGGCAATGCTTCATT..... 2481
818 LeuAsnLysGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAspGly 834
2482AATCTAAGCAGCACCGCCGACAA..... 2505
834 yProValAsnIleSerThrAlaLeuSerLeuAsnSerArgProAspG 851
2506AACGCACTGACG 2520
851 lValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerTrpAsn 867
2521 CTTCCTCCGCG...AACGCTAAGCAACGTAAGC...CATTCGCGACTCA 2564
868 LeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeuSe 884
2565 CGGTAAATGCTCCCTAGCCGATAG..... 2589
884 rGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGlyGln 901
2590CGAGTA 2595
901 LysLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSerLeu 917
2596 TTCATTTTGAAGACGCGCTTACCGGACAAATCAGCGCGGCAAGA 2645
918 PheAsnGlyTyrArgAsnIleTrpSerGlySerLeuAsn...AlaProAs 933
2646 TACGCGATTAACCTTAAAGACAGCAATGAGCGCTGCCGTGAGGACGG 2695
933 PalatThrValSerMetThrAspThrGlnThrSerMetAsnGlyAsnSer 950
2696 AATTAGCAATTTAAACTTTCACACGCGCACATTCATCAATTCGCC 2745
950 hValaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsnGlyGly 966
2746 TATGCGCAGATGGCGGCGGCAACCGCAGTCGACAGATGGCC 2795
966 966
2796 GCGCGCGCTTGGCGCGCTTCGCCGCTTCCTATTAATCCGTTACCGCC 2845
966 966
2846 CAACTTGGTAGAATCCCGTTTCAACAGCTGACGTAACGCAATTTG 2895
967ThrSerProPheThrThrLeuThrThrAspAsnLeuAsp 979
2896 AACGCTGAGGCAACATTCGCTTTATGTGCGAACTTCGCTACGCGAG 2945
980 AlaValGlnSerAlaPheValMetArgThrAspLeu.....AsnLysAl 994
2946 CGCAAAATTTGAAGCTGGCGGAAAGTTCGAAAGGCACTTACACCTTGGCG 2995
994 aspLysLeuValIleAsnLysSerAla..... 1003
2996 TCAACATTAACGGCAACGACACTCAAGCTCGAACAATTAAGAGTAGT 3045
1003 1003
3046 GAAGCAAAAGACAAACCGCTGTCCGAAACCTTAATTTACACCTGCA 3095
1004 ThrGlyHisAspAsn.....SerIleTrpValAsnPhe..... 1014
3096 AAACGAACAGCTGATCGCGCGGTGCGCTTACCAATCATATCGCAAG 3145
1014 1014
3146 ACGCGAGTTCGCGCTCATTAATCCGTTCAAGACAAAGAGCTTTCGAC 3195


```

||| ..... |||
27 ThrCysLysSerHisArg.....ArgLeuSe 35
72 GCCCTGCTACTATGACCATATGCTGCTGGCATTCCTCCCAACCT 121
| ..... ||| ..... |||
35 rArgArgValIleLeuThrSerValAlaLeuSerLeuSerLeuIat 52
122 GG.....GCGGACACACTTATTTCCGCATCACTACCAATACTAT 162
||| ..... ||| ..... |||
52 rPProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68
163 GCGGACTTTGGCGAAATAAAGCAAGTTGCACTCGGCGCGAAGATAT 212
||||| ..... ||| ..... |||
69 ArgAspPheAlaGluAsnLysGlyInPheThrProGlyThrThrAsnI 85
213 TGAGTTTACAAACAAAAAGGAGTGTGCGCAATCAATCAACAAAG 262
| ..... ||| ..... |||
85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101
263 CCGCGATGATGATTTTCTGTGTGCGGTAC..... 297
||||| ..... ||| ..... |||
101 lApProMetAlaAspPheSerSerAlaThrIleThrIleThrGlySerLeuPro 117
298 ...GCGTGGCGGATGTGTGGGATCAATATATTTGTGAGCTGGCACA 344
||| ..... ||| ..... |||
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValValThrAlaLysH 134
345 TAAACGCGGCTATACAAACGTTGTTGGTGGGAGAGAAATCCCG 394
| ..... ||| ..... |||
134 sValSerGlySerAspThrMetSerPheGly..... 144
395 ATCAACATCGTTTACTTATAAATGTGAAACGGATATATTAAGCA 444
145 ..... TyraLysAsnThrTyrThrAla 152
445 GGGACTAAAGCCATCCTTATGCGGCGATTAATCATATGCCGTTTGA 494
| ..... ||| ..... |||
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
495 TAAATTTGCACAGATGCAGAAACCTGTTGAAATGAC..... 531
||||| ..... ||| ..... |||
169 rLysLeuValThrGluValAlaProAlaGluValSerAspIleGlyAlaV 186
532 ..... AGTTATATGATGGGCGGAATATATGATCAAAATATATTAC 573
| ..... ||| ..... |||
186 alSerGlyAlaTyrGlnAlaGly.....GlyArgPhe 196
574 CCTGACCGTGTCTATTGGGCGAGCAGCAATATTTGGCGATGATGA 623
| ..... ||| ..... |||
197 ThrGluPheTyrArgLeuGlyGlyMetGlnTyrValLys..... 210
624 AGATGAGCCCATTAACCGGAAAGTTCATTCATATTGCAAGTCGATAT 673
||||| ..... ||| ..... |||
211 AspLysAsnGlyAsnArgThrGlnValTyr.....ThrAsnGlyG 224
674 GTTGGCTGTTGGTGCATACTTTGCACAAATGATGACGAGTGGCG 723
| ..... ||| ..... |||
224 lPheLeuValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGly 240
724 ..... ACAGTCACTTAGTAGTGAATAAATTAACATTAACCCATA 764
||||| ..... ||| ..... |||
241 GlnMetIleThrAlaGlnThrGly...AspIlePheLeuIleProProAs 256
765 TGGTTTTTACCAACAGAGGCTCATTTGGGAGCGGCTCAACCATGT 814
||| ..... ||| ..... |||
256 nGlyProLeuAlaAsnTyrLeuAsnMetGlyAspSerGlySerProLeuP 273
815 TTATGTATGATGCCCAAAAGCAAAAGTGTATTAATTAAGGGTATTGCA 864
||| ..... ||| ..... |||
273 heaIaTyrAspSerLeuGlnLysLysTyrValLeuIleGlyValLeuSer 289
865 ACGGCAACCCCTATATAGAAAGCAATGCGTCCAGCTGTTGCTAA 914
| ..... ||| ..... |||

```

```

290 SerGlyThrAsnTyr.....GlyAsnAsnThrValValThrThrG 303
915 AGATTGTTCTATGATGAATCTTTGCTGGAGATACCATCATGATTTCT 964
| ..... ||| ..... |||
303 nAspPheLeu..... 306
965 ACGAACCACGTCAAAMTGGGAATACTTTTAACGACGATAATATGCG 1014
| ..... ||| ..... |||
307 .....GlyGlnGlnProGlnAsnAsp..... 313
1015 ACAGAAATAATCATGCGCAACATGAAACACATTCCTCTGCTATATGATT 1064
313 ..... 313
1065 AAAAACAGAACCGTCAATGTTTATGTTTCTTATTCGAGACAGCA 1114
| ..... ||| ..... |||
314 .....PheAspLysThrIleAlaTyrThrSerG 323
1115 GAGAACCTGTT.....TATCATGCTGACAGTGGTGTCAAC... 1149
||||| ..... ||| ..... |||
323 lGlyGluGlyValLeuGlnThrPlyAspPheAlaAlaAsnGlyThrGlyThr 339
1150 .....AGTTATCGAACCGA..... 1164
340 LeuThrGlnGlyAsnThrThrTyrGlyTyrAlaThrPlyGluArgLysLe 356
1165 .....CTGATATATGAGAAATATTTCTTTATTTGACAGAGAAAGCG 1210
||||| ..... ||| ..... |||
356 uIleLeuAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyG 372
1211 AATTGATCTTACCAGCAACATCAATCAAGTGTGCTGAGAGATTATATTC 1260
| ..... ||| ..... |||
372 lValValIleuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPhe 388
1261 CAAGAGATTTTACGCTGCTGCTGCTGAAATTAACGAACCTGGCAAGGCG 1310
||||| ..... ||| ..... |||
389 AlaGlyAspTyrAlaGlyValSerAlaLeuAsnGlyGlnThrThrPheGlyG 405
1311 GGGCTTCATATCATGTAAGACAGTACCGTTACTTTGAAAGTAAACGGCG 1360
| ..... ||| ..... |||
405 yGlyIleIleThrAspLysGlyThrHisValIleuThrGlnValAsnGly 422
1361 TGGCAACGACCGCTGTCGCAAAATCGGCAAGGACCGCTGACGTTCAA 1410
||||| ..... ||| ..... |||
422 alaIleGlyAspAsnLeuHisLysThrGlyGlyGlyThrLeuThrValAsn 438
1411 GCCAAAGGGGAAACCAAGGCTCGATTCAGCGTGGCGCAGCTACAGTCAT 1460
| ..... ||| ..... |||
439 GlyThrGlyValAlaAsnAlaGlyGlyLeuLysValGlyAspGlyThrValI 455
1461 TTTGGATCGAGGAGAGCATTAAGGCAAAACCAAGCCTTTTGTGAAA 1510
||||| ..... ||| ..... |||
455 eleuAsnGlnGlnAlaAspArgAspGlyLysValGlnAlaPheSerSerV 472
1511 TGGCTTGGTCAAGCGGAGGCTGACGCTGCAACTGAATGCGGATTAATCAG 1560
| ..... ||| ..... |||
472 alGlyIleAlaSerGlyArgProThrValValLeuSerAspSerGlnGln 488
1561 TTCACCCCGCAAACTATTTGGCTTTGGCGGCGAGCTTTGATTT 1610
||||| ..... ||| ..... |||
488 ValAsnProAspAsnIleSerThrPheGlyTyrGlyGlyValThrLeuGlnI 505
1611 AAACGGGATTCGCTTGTGCTCCACCGATTCGCAAAATACGATGAAGGG 1660
||||| ..... ||| ..... |||
505 uAsnGlyAsnAsnLeuThrThrPheThrArgLeuGlnAlaAlaAspTyrGlyA 522
1661 CGATGATTTGTACACACATCAAGACAAAGATTCACCGCTTACATTT... 1707
| ..... ||| ..... |||
522 lAlaIleIleThrAsn...AsnSerGlyLysLysSerThrValThrLeuAsp 537
1708 .....ACAGCATAAAGATTTGCTACAAACCGGCAATACAAACAG 1748
| ..... ||| ..... |||
538 LeuGlnThrLeuLysAlaSerAspIleAsnValProValAsnThrValSe 554

```


1749 CTTC.....GATACCAA. 1761
1762
554 rtlephgilyarglyalaproglyaspleutyryaspsersert 571
1762AAGAAATTCCTACACGGTTGGTTGGC 1791
571 hrlysglnryrphelleuleylasersertyserpropheser 587
1792 GAG.....AAGATTCAGCCAA 1808
588 Aspleuansanserseryaltrpqlansvalglylaspshiasnly 604
1809 AAC..... 1812
604 salalleasplhrVallysglnlyslleglualaserserglnpror 621
1813AACGGCGCTCAACCTTGTTCACCAAGCC 1842
621 yrmetyrhlsglylnleuansnlyasmetaspvalasnileprogin 637
1843 GCCGCAAGACCGCACCTCTGCTTCGCGGCAACAATTTA..AA 1889
638 leuserglyaspyalaleualaleuaspolyserValasmeuprogl 654
1890 CGGCAATCAGCAACAAGCGCAACTGTTTTCAGCGGCAACCA 1939
654 uglyserlthryllyserglythrleuilephenglnlyhlsprlv 671
1940 CACGCGCGCTACATCATTTAAAGCACATTCGTGCAAAAAGCGC 1989
671 alllehsalaglythrThrserserglnser..... 683
1990 ATTCTCGCGGGAATCGTGTGGACACAGCTGATCAGCGCAT 2039
684AspTrpelnthrarglnph 690
2040 TAAACCGGAAATTCGAATTAAAGGCGACGCGGTTCGCCGA 2089
690 ethrleuileuylsleuylsleuaspalalathrPhehlsleuserlrga 707
2090 ATGTCGCAAAAGTGAAGCGATTGGCAATTTGACCAATCAGCCCAAGCA 2139
707 sn...glylsmetcnglnlyaspilasnlatrhansglyserthrVal 722
2140 GTTTTGGTGTGCGACCCGATCAAGCCACACATCTGTACAGCTGGA 2189
723 ileuengly.....serSerarvalphethrtrspargserAs 735
2190 CTGGACGGGT...CTGACAAATGTCGAAAAAACCAATTACGACGATA 2236
735 pelythrnglyasnalaivalserSerValgluglyseralathralthrt 752
2237 AAGTATTCCTTCATTGACTAGACGACATCAGCGGCATGTGCATCT 2286
752 hrVal.....glyaspGlnserAspTyrserglyasnValthrleu 765
2287 GCCGATCAGCTCATTTAAATCTCACAGGCTTGCCACATCAAGCGCA 2336
766 gluasnlyserSerleuenglnle.....Metgluatgphethrnglygl 780
2337 TCTTAGTCGCAAAATGCGATACAGTTATACATCAGCCAAAGCGCACCC 2386
780 yllleglualatryaspsertThrValserValthrSerGlnsnalaValp 797
2387 AAAACGCAACCTTACCTCTGCGGCATGCCCAAGCAACAATTATACA 2436
797 heasprargValglySerpheValasnSerSerleuthrleuenglyllysl 813
2437 GCCACATTAAGCGCAACATCGGCTTCGCAATGCTTCATTAACCT 2486
814 Alalysleu.....Thralaglnserglyllepheserthrclyal 827

2487 AAGCAGCACCGCGCTACAAAAGCGAGTCGATCGCTTCGCGCAACGCTA 2536
827 aValasp...VallysgluasnalaSerleuthrleuthrleuethrlymetros 843
2537 AGCGAATCCTAAGCCAT.....TCCGACCTCAACGCT 2568
843 erlathlnlysglnlytyrtyrSerprovallleserthrthrlugly 859
2569 AATGTCTCCCTAGCGATAGGCACTATTCATTTGAACAGCGCGCTT 2618
860leasnleucluaspsasnlaSerPne.....SerValysAs 872
2619 TACCGCAACAATCAGCGCGCGCAAGGATTCGCAATTACACTTAAGACA 2668
872 nmetglytyrleuserSer..... 878
2669 GCGAATGACGCTCGCTCAGCAGCGAATTCGCAATTTAACCTTGAC 2718
879AspIlehsalaglythrThrAlalathrIleasnleucl 892
2719 AAGCGCACCATTCACCTCAATTCGCGCTATCGCACGATCGCGAGGCG 2768
893 Asp.....SeraspAl 896
2769 GCAAAACCGGAGTGCAGACAGATCGCGCGCGCTTG...CGCGCT 2815
896 aasplalaglylys...ThrAsperProleupheserSerleuMetlysg 912
2816 CGCGCGCTTCCTCAATTA.....TCCGTTACACCGCACTTCGCTGAGA 2859
912 lytyrAsnalavalaleuarglyserIlethrcllyalaglnserthrVal 928
2860 TCCGTTTCACACGCTGACGGTAAAC.....GCCA 2891
929 AsnmetlleasnlatleuThrPtyrseraspglyllyserclualalalyal 945
2892 ATTGACGGTCAGGAGACATTCGCTTATGTCGAACTTCGCGCTAC 2941
945 aleuylsAlalysglyserargllegluleuclayaspolylyshlsPhea 962
2942 GCACGCAAAATTCGACCTGCGGAAAGTTCGAGAGCACTTACACTTG 2991
962 latrhleuGlnVallysgluuenserAlaspasnthrThrPheleuMet 978
2992 GCGGTCAACATACCGGCAACGACCTGCAAGCTGCAACAATTCAGGT 3041
979 HlsThrAsnAsn.....SerargAlaspelnleuasnva 990
3042 AGTGGAA.....GGAAGACAAACAACCGCTGTCGAAAACCTTA 3082
990 lThrAsplysleuserglyserAsnAsnserValleuValAspheleuA 1007
3083 ATTTCAACCTGCAAAAAGCAACAGCATCGATGCGCGCTGTACCA 3132
1007 sn..... 1007
3133 CTCATCCGCAAAAGCGGAGTTCGCGCTGATATCCGCTCAAGACA 3182
1008LysProAlaserglume 1013
3183 AGAGCTTCGCAACAATCTGCGCAAGGCAAGGCCAAAAAACAGCGGAAA 3232
1013 tseryalthr..... 1016
3233 AAGACAGCGGCAAAAGCTTCAGCGCTGATGCGCGCGGAGATGC 3282
1017leuIlethrAlalProlysglySer 1024
3283 GTCGAAAGACAGAAAGGTTCCGACACCGCGCGGAGCGAGCGGGA 3332
1025 AspglyllysthrPheThr.....Alaglythrngl 1034
3333 AATGTGCGCATTTATGCAAGCGGAGAAAGAAAAACGGGTGACGCGG 3382

```

1034 nclnleeglypserasnvalthrprovallelserthrclulyszhra 1051
3383 ATAAAGACGACGCGCTGGCGAAAGCGGCAAGCGCAACCGCGCGCT 3432
1051 spaspalatrhr.....lystrpvalleuthrlytyrlygln 1062
3433 ACCACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTGGCGCAACT 3482
1063 thrthrallaespalaglyalaserlyalalalalysasp..... 1075
3483 GCAACCCCAACCGCAGCCCAACCGCAGCGCATGATCAGCGCTTAA 3532
1076 .....PheM 1077
3533 CCAATACGCGTTGATGATTTTCGCGCACGCTCAACAGCGTTTCCG 3582
1077 etalaserglytyrlyserphenleuthrclulalasnleuasnlys 1093
3583 GTACAGACGCAATTAGACCGCGTATTTGCGAAGACCGCGCAAC... 3627
1094 Argmetclyspleu.....Argaspthrclnglyaspal 1105
3628 .GCCGTTTGGACAAGCGGATCCGGGACACCAACACTACCGTTGCAAG 3676
1105 aglyvaltrpalaargllemetasnlythr.....glyseralaa 1119
3677 ATTTCCGCGCTACCGCGCAACAGCGCTGCGCAATTCGTTATGAG 3726
1119 sp...glyasptyrserasptsttythrhisvalclnleelyalasp 1134
3727 AAA.....AACCTGCGCAGCGCGCGCGCGCAT 3755
1135 Arglyhisclnleuaspglyvalaspleuphethrclgialaleuth 1151
3756 CCGTGTTCGACACCGGACCGCAAAACACTTC.....GACGACG 3796
1151 rtyrthrasptserasnalaaserthrhisalapheserglylysasnlys 1168
3797 GCATCGCAACTCGCGACGCGTGGCCACGCGCGCTTTCGCGCAATAC 3846
1168 ervalclgly..... 1171
3847 GGCATGACAGGTTTACATGCGCATCAGCGCGCGCGCTTTAGCAG 3896
1172 .....glyleuthrlyalaseralaleuphasns 1181
3897 CGGACGCTTTGACAGCGCATCGAGGAAATCGCGCGCGCGTGC 3946
1181 rglalalyrphenaspleuilegly.....lystryleu 1193
3947 ATTACGCGCATTCAGCAGCATACCGCGCGCTTTCGCGCATC 3996
1193 lshlaspasncln.....Hlthrallaasnphenalaserleuglythr 1207
3997 GAACGCGCATCGCGCAACGCGTATTTGCAAAAGCGATTCACG 4046
1208 lysasptyrserthrhisertprtyrlyalaglyalaglyalaglythr 1224
4047 CTAC.....GAAACGCTCA 4060
1224 gtyrthlserthrlysglusertrpvalcluproglinleuvalt 1241
4061 AATATGCGACCGCGCTTCATTCACCGCTACCGCGCGCATTAAG 4110
1241 yrclyservalsercllysalaphesertrpglualaargglymetala 1257
4111 GCAGATTTCATTCACCGCGCAACACATTCATTCACCGCTTATT 4160
1258 .....LeuserMetlyasplyasptyrasnproleuileglyalagly 1272
4161 GAGCGTGTCTATACCGATCCGCGCTTCGCGC.....AAAGTCGAA 4201

```

```

1272 rglvalaspyalglyarglapheserglyaspsptprlyllethr 1289
4202 CACGC.....GTCAATACCGCGCA 4221
1289 laargalaglyleuglytyrclnphaspleuclnleuasnlyglutthr 1305
4222 TTGGCTCAGGATTTTCGCAAAACCGCGATGCGCAATGCG..... 4260
1306 valleuclnaspalaserglyglutlysargprhegluglyclulysasps 1322
4261 .....GGCGTAAACCGCGCAATCAAA 4281
1322 raArgMetleuMethrvalclglymetasnlaaglyleu 1335

seq_name: p1r2:S61319
seq_documentation_block:
  Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (f
  C:Species: Neisseria meningitidis
  A:Variety: NG17
  C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
  C:Accession: S61319
  R:Lombolt, H.; Poulsen, K.; Mogens, K.
  Mol. Microbiol. 15, 495-506, 1995
  A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
  A:Reference number: S61314; MUID:95302961
  A:Accession: S61319
  A:Status: preliminary; nucleic acid sequence not shown
  A:Molecule type: DNA
  A:Residues: 1-508 <LOW>
  A:Cross-references: EMBL:X82483; NID:g732881; PIDN:CAAS7866.1; PID:g732882
  C:Superfamily: Iga-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloproteinase

alignment_scores:
  Quality: 890.00      Length: 563
  Ratio: 2.500         Gaps: 21
  Percent Similarity: 63.233      Percent Identity: 38.721

alignment_block:
  US-09-303-518D-649 x S61319 ..
  Align seg 1/1 to: S61319 from: 1 to: 508

202 GCGAAGATATTGAGCTTTACACAAAAAGGAGCTTGCGCAATC 251
   ||| |||:||||| |||:|||||: |||:|||||: |||:|||||:
   1 Alatrhaspleuservallysasnlysglnclglnasnleeglyasnal 17
252 AATGACAAAGCCCGATGATGATTTTCGTGGTGCAGCGGTAC...G 298
   |||:|||||:||||| |||:|||||: |||:|||||: |||:|||||:
   17 aleuserasnvalPrometilleaspheservalalaspalasnlysa 34
299 GCGTGGCGCATTTGCGCGCATCAATATATTGTAGCGTGCACAT... 345
   |||:|||||:||||| |||:|||||: |||:|||||: |||:|||||:
   34 rgllealatrvalalaspProglntyrlyalalaserlyalalala 50
346 .....AACGCGCGCATTAACAACGTTGATT 371
51 lysalaglyalalsthrphenetyrtyrclnglylnlyasnclnlysnas 67
372 TGGTCGCAAGAGAAATCCGATCAACATCGTTTACTTATTAATAATG 421
   |||:|||||: |||:|||||: |||:|||||: |||:|||||:
   67 ryalalaspysgluasn.....glutylarvaly 78
422 TGAACGGAATATTATTAACAGGACGACTAAGGCCATCTTATGGCGC 471
   |||:|||||:||||| |||:|||||: |||:|||||: |||:|||||:
   78 algluglnasnasnlyrclupro.....Hltsalatrpgly 90
472 .....GATTATCATATGCGCGCTTGCATAA 497
   |||:|||||: |||:|||||: |||:|||||: |||:|||||:
   91 Alaserasnleuglyargleuglnasptyrasnmetalaargphasnly 107

```


Align seg 1/1 to: S61315 from: 1 to: 496

```

202 GCGAAGATATTTGAGGTTTACAGAAAAAGGAGAGTTGGTCGGCAATC 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 AlatrhispleuSerValLysAsnLysGlnGlnAsnIleGlysnal 17
252 AATGCAAAAAGCCCCGATGATGATTTTCTGTGTCGTCGCTAAC...G 298
    : : : : : : : : : : : : : : : : : : : : : : : : :
17 aleuSerAsnValProMetIleAspPheSerValAlaAspValAsnArg 34
299 GCGTGGCGGCGATTTGGTGGCGATCATATATTTGTGAGCTGGCAATAC 348
    : : : : : : : : : : : : : : : : : : : : : : : : :
34 rghIleuThrValIleAspProGlnIleValAlaSerValLysHisVal 50
349 GCGGCTATACAGACGTTGATTTTGGTGGCGGCAAGAAATCCCAT... 396
    ||| : : : : : : : : : : : : : : : : : : : : : :
51 LysGlyAspGlnIleSerIleGlyGlnHisAsnGlnLysIleuAspVal 67
397 ...CAACATCGTTTACTATATAAATTTGAAACGCAATTAATTAAGA 442
    : : : : : : : : : : : : : : : : : : : : : : : : :
67 lSerAsnAspLysAsnGlnIleIleArgSerValAlaGlnAsnAspTyrGln 84
443 CAGGCACTAAAGCCATCCTTATGGCGGC...GATTAT 477
    : : : : : : : : : : : : : : : : : : : : : : : : :
84 roAsnLysAsnIlePheHis...HisGlyAsnGlnGlyArgLeuGlnAspTyr 99
478 CATATCGCGGCTTGCATTAATTTGTACAGATGACAGAACCTGTGAAT 527
    : : : : : : : : : : : : : : : : : : : : : : : : :
100 AsnMetLysArgLeuAsnLysPheValThrGlnValAlaProIleAlaIle 116
528 GACCACTATATGATGGG...CGAATATATCCATCAATTAATTAACC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 oThrSerAlaGlyLysGlyValGlnIleThrTyrLysAspLysAsnArgPheS 133
575 CTACCGCTGTCTATTTGGGAGCGAGCGCAATATTTGGCATGTGATGA 624
    : : : : : : : : : : : : : : : : : : : : : : : : :
133 erGlnPheValArgValGlnIleGlyThrGlnPhe... 144
625 GATGAGCCCAATACCGCGAAGTTCATATCATTTATGCAAGTGCATTC 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ...GltTyrAsnSerIleTyrAsnMetThrGlnIleuSerArgAlaTyrAr 160
675 TTGGCTCGTGTGGCAATACCTT...GCACAAAATG 709
    : : : : : : : : : : : : : : : : : : : : : : : : :
160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValIleThrSerAsnL 177
710 GATCAGAGTGTGGCACAGTCAACTAGTAGTGAATAAATTAACATAGC 759
    : : : : : : : : : : : : : : : : : : : : : : : : :
177 euAsnGlnIleGlyLeuIleGlyPheGlyAspAsnSerLysHisIleSer 193
760 CCA...TATGTTTTTACCAACAGAGG 785
    ||| : : : : : : : : : : : : : : : : : : : : : :
194 ProGlnLysLeuLysGlnValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210
786 CTCATTTGGCGACAGTGGCTCCACCAATTTATCTATGATGCGCAAAAGC 835
    : : : : : : : : : : : : : : : : : : : : : : : : :
210 aValIleuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGul 227
836 AAAAGTGGTATTAATAGGGTATTTGCCAAAGCGCAACCCCTATATAGA 885
    : : : : : : : : : : : : : : : : : : : : : : : : :
227 ysArgTTrValPheLeuGlnIleValTyrAspTyrTrpAlaGlyTyr... 241
886 AAAAGCAATGCTTCACAGCTGTTCGTAAAGATGG...TTTCATGATGA 932
    : : : : : : : : : : : : : : : : : : : : : : : : :
242 GlnLysAsnSerTrpGln...GltTyrAsnIleIleTyrLysLys 254
933 AATCTTCTCTGAGATACCATTCAGTATTTCTAGAACACCACTCAAAATG 982
    : : : : : : : : : : : : : : : : : : : : : : : : :
254 sGlnPheAlaAsp...GlnIleLysGlnArg 263
983 GGAATACTCTTTAAGAGATATAATGACAGAAATAATCAATGCC 1032
    ||||| ||||| ||| : : :

```

```

264 .....AspAsnAlaGlyThrIleLysGlyTyrGly 273
1033 AAACATGAACAAATTCCTGCCAATAGATTAAAAACAGAACCTTCA 1082
    : : : : : : : : : : : : : : : : : : : : : : : : :
274 GlnHisThrIlePlyStrHis...ThrIleThr... 282
1083 ATTGTTAATGTTCTTATTCGAGACGACGACAGAACACCTGTTATCATG 1132
    ||| : : : : : : : : : : : : : : : : : : : : : :
283 .....AsnSerHisIleGlySerThrIleValAlaGlyLeuAlaGlyAsnG 297
1133 CTCGAGGCGGTGCACACAGTTATTCGACCCAGACTGATTAATGAGAAAT 1182
    ||| : : : : : : : : : : : : : : : : : : : : : :
297 LysArgGlyAla...AsnAsnGlyGlnAsn 305
1183 ATTTCCTTATTTAGCAGAGAAAGCGAAATGATGATTCATCCAGCAAT 1232
    : : : : : : : : : : : : : : : : : : : : : : : : :
306 ValThrPhe...GlnAspAsnGlyThrLeuValLeuAspGlnAsnIle 320
1233 CAATCAAGGCTGTGGAGGATTAATTTCCAAAGAGATTTTACGCTCCG 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 eAsnGlnIleValGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337
1283 CTGAATAATTAAGAA...ACTTGGCAAGCGCGCGGCTCATATCACTGA 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 LysAlaAsnAsnAspIleThrThrPheGlyAlaGlyIleAspValAlaAsp 353
1330 GACAGTACGCTTACTTGAAGTAAACGGCGTGGCAACAGCCGCTGTC 1379
    : : : : : : : : : : : : : : : : : : : : : : : : :
354 GlyLysLysValValThrPheGlnValLysAsnProAsnGlyAspArgLeuAl 370
1380 CAAATTCGCGCAAGGACGCTGCACGCTTCAACGCAAGGCAAAACGAG 1429
    : : : : : : : : : : : : : : : : : : : : : : : : :
370 aLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnG 387
1430 GCTCGATCAGCGCTGGCGGACGAGTACATCATTTGATGATCAGCAGCAG 1479
    ||| : : : : : : : : : : : : : : : : : : : : : :
387 LysIleuLysValGlyAspArgLysThrValIleLeuAsnGlnIleAlaAsp 403
1480 GATTAAGGCAAAAAACAACCTTGTAGTCAATCGCTTGTGTCAGCGAG 1529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420
1530 GGGTACGCTGTCAATGATCCGATATACGTTCAACCCGACAAACGCT 1579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 gGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 437
1580 ATTTCGCTTTCGCGCGGACGCTTGTGATTAACAGGCGCATTCGCTTCG 1629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 yThrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCACCGGATTCAAATACCGATGAGGCGGCGATGTGTCACACGCA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAs 470
1680 TCAACACAAAGATCCACCGTTACCATTAACAGCAATTAATGATGCTA 1729
    ||| : : : : : : : : : : : : : : : : : : : : : :
470 nThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIle 486
1730 CAACCGCAATTAACACAGCTTG 1752
    ||| : : : : : : : : : : : : : : : : : : : : : :
486 hTrAlaProGlnAsnLeuSerVal 493

```

seq_name: p1r2:S61328

seq_documentation_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr

C;Species: Neisseria meningitidis

A;Variety: NG44/76

C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

R;Lombolt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A;Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss


```

1380 CAAATCGGCAAGGCGCTGCACCTTCAAGCCAAAGGGAACCAAG 1429
      |||||
370 alysiilegylslythrlleuulileasnlythrllyalasnlns 387
      |||||
1430 GCTCGATTCAGCGTGGCGGACGCTACAGTCAATTTGGATCAGCAGCAGC 1479
      |||||
387 lyclineuysvalglyaspilgylthrlleleuasnlnlyalasp 403
      |||||
1480 GATAAGGCAAAACACCTTTAGTGAATCGCTGGTGGAGGAG 1529
      |||||
404 Serasnlnlyvalglinalpheserelnvalglilevalsergllyar 420
      |||||
1530 GGCTAGCGTGCACACTGATGCGATATCACTTCAACCCGCAAACTCT 1579
      |||||
420 ggllythrlleuvalleuasnserferasnlnleasnproaspasleut 437
      |||||
1580 AATTGGCGCTTTCGGCGGACGCTTGGATTTAACGGCGATTCGGCTTTCG 1629
      |||||
437 yphelglyphearglygllyarqleuaspalaasnlyasnspleunhr 453
      |||||
1630 TTCACCGGATTCACAAATACCGATGAAGGCGGATGATTCACACAGAA 1679
      |||||
454 pheelhlnstleargasnvalaspilgylalarglilevalasnlnas 470
      |||||
1680 TCAAGACAAAGAAATCCACCTTACCTTACGAGCATTAAGATATTGCTA 1729
      |||||
470 nhrcllysthrserthrllethrlleuthrlgly...lyserleuilef 486
      |||||
1730 CAACCGGCAATACACACGCTTGATAGC 1758
      |||||
486 hrasnproaspasleuSerValHisSer 495

```

seq_name: p1r2:S61325

```

seq documentation block:
  Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C:Species: Neisseria meningitidis
A:Variate: SM1894
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61325
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseri
A:Reference number: S61314; MIMD:95302961
A:Accession: S61325
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82476; NID:g732867; PIDN:CA57859.1; PID:g732868
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
  Quality: 870.00      Length: 543
  Ratio: 2.403        Gaps: 19
  Percent Similarity: 66.667      Percent Identity: 38.674

```

alignment_block:

US-09-303-518D-649 x S61325 ..

```

Align seg 1/1 to: S61325 from: 1 to: 496
202 GCGAAGATATGAGGTTTACAAAGGAGGAGTTGTGCGCAATC 251
      |||
1 AlatrhraspleuSerVallylsnlysglnlglinalasnllleglyasnAl 17
      |||
252 AATGACAAAGCCCGATGATGATTTCTGTGTGTCGCGTAAC...G 298
      |||
17 aleuSerasnvalPrometileaspPheSerValalaspValasnArgA 34
      |||
299 GCGTGGCGGATTTGGTGGCGATCAATATTTGTGAGCGTGGCAATAC 348
      |||

```

```

34 rglthrlleuThValIleasnproglIntyrlalalaserVallysnlval 50
      |||
349 GCGGCTATTAACACCTTGATTTGTGCGGAGGAAGAAATCCGAT... 396
      |||
51 lylslyaspilgylleuSerlythrllythrlshlshasnlnlysnleuaspva 67
      |||
397 ...CAACATGCTTTTACTTAAATGTGAACGGAATATTAATTAAG 442
      |||
67 lserasnspilglnasnlnlythrlarqSerValalaglinalasnsplyrclup 84
      |||
443 CAGGACTAAGAGCCATCCTTATGGCGGC...GATTAT 477
      |||
84 roasnlyasntrpHis...HslglasnlnlgllyarqleuGlualsptry 99
      |||
478 CATATGCCCGCTTGTGCAATAATTTGTACAGATGACAGACCTGTGAAT 527
      |||
100 AsnmetlalalargleuasnlysnPhevalthrlglualalapollelealpr 116
      |||
528 GACCACTTATATGATGGG...CGGAATATATGCAATAATTAATAC 574
      |||
116 othrseralagllygllyalgllythrllytrlysaspIysasntrpHis 133
      |||
575 CTGACCGCTGTGATTTGGCGGAGGACGCAATATTTGCGATCTGATGAA 624
      |||
133 ergluphevalargvalgllyalagllythrlgnPhe... 144
      |||
625 GATGAGCCCAATATACCGGAAAGTTCAATATCAATATGCAAGTGCATTC 674
      |||
145 ...GlutyrasnserarqlyrlyasnmetthrlgluleuSerarqlyar 160
      |||
675 TTGGCTGCTTGGTGCATATACCTT...GCACAAATG 709
      |||
160 gtyralalilealagllythrlprotyrlnaspvalasnvallnrserasn 177
      |||
710 GATGAGTGTGTGCACACACTTACTAGTACGAAATAATTAACATAGC 759
      |||
177 euaSnlnlgllygllyleuIleagllyPhegllyAspasnserlyshlshSer 193
      |||
760 CCA...TATGCTTTTACCAACGAGG 785
      |||
194 ProglulysleuysglualleuSerelnasnvalaleuthrlnasnlyrAl 210
      |||
786 CTCATTTGGGAGACGCTGCTCACCAATGTTATTCATGATGCCCAAAAC 835
      |||
210 aValleuGllyaspserlyserProleuphealalaryaspIyslnlglul 227
      |||
836 AAAAGTGTTAATTAATGAGGATTTGCAAAAGGCAACCCCTATATAGA 885
      |||
227 ysarqtrpvalPheleuGllyalalatyrsPtyrtrpAlagllyr... 241
      |||
886 AAAAGCAATGCTTCCAGCTGCTGCTGTAAGATTTG...TTCTATGATGA 932
      |||
242 GlulysasnsertrpGln...Glutyrasnlnlelyrlyslly 254
      |||
933 AATCTTCTGAGATACCCATTCAGTATTCAGAACACGTCATAATG 982
      |||
254 seluphealasp...GlulileylglalnArg. 263
      |||
983 GGAATACTCTTTTACGACGATATATATGCGACAGAAACCAATGCC 1032
      |||
264 ...AspasnlnleagllythrllelysglytyrGly 273
      |||
1033 AATCATGAACACCAATTCCTCCATATATGATTAATAAACGACGACCTTCA 1082
      |||
274 GlulshlshstrpIysThr...ThrlglThr... 282
      |||
1083 AATTGTTAATGTTCTTATACCGACAGACAGCAAGAACCTTTATATG 1132
      |||
283 ...AsnserhlnstlegllySerthrlalavalalargleuAlalysnng 297
      |||
1133 CTGCAAGTGTGTCAACAGTTATGACCCAGACCTGCAATATGAGAAAT 1182
      |||
297 lualrglyala...Asnasnlyglalnasn 305
      |||

```



```

1183 ATTTCCTTATTTGACGAAAGGAAAGCCGCAATTGATCTATTACCAGCAACAT 1232
1184 .....||| |||.....||| |||.....||| |||.....|||
306 ValThrPhe..... GluAsnAsnGlyThrLeuValLeuAspGlnAsnIle 320
1233 CAATCAAGCTGCTGAGAGATTATATTTCCTCAAGAGGATTTACGGCTGCGC 1282
1234 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
320 eAsnGlnGlyIleGlyGlyLeuPhePheGlyAspPyrThrValIysG 337
1283 CTGAAATAAAGGAA...ACTTGGCAAGCGCGCGCGCTTCATATCAGTGAA 1329
1284 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
337 LylLeuAsnAsnAspIleThrTrpLeuGlyIleAspValAlaIleAsp 353
1330 GACAGCTACCTTACTTGGAAAGTAAACGGCTGGCAACGACCGCTGTC 1379
1331 ..: |||.....||| |||.....||| |||.....||| |||.....|||
354 GlyIysLysValValTrpGlnValLysAsnProAsnGlnAspArgLeuAl 370
1380 CAATATGGGCAAAAGGACGCTGCACGTTCAAGCCAAAGGGGAAACCAAG 1429
1381 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
370 aLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnG 387
1430 GCTGATCATCAGCGTGGCGAGCGTACAGCATTTGGATTCAGCAGCAGAC 1479
1431 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
387 LysIleLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaIAsp 403
1480 GATAAAGGCAAAAACAACGCTTTAGTAAATCGGCTTGGTCACGGCGAG 1529
1481 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
404 SerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420
1530 GGGTACGCTGCACACTGAATGCGCGTATTCAGTTCACCCGACAACTCT 1579
1531 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
420 gArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsnLeuThr 437
1580 ATTTCGGCTTTCGGCGGCGAGCGTTTGATTTAAACGGCGCAATGGCTTTCG 1629
1581 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
437 YrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCACCGCTTTCAAATATCCGATGGAAGGGCGCATGTTGTCAACCAAA 1679
1631 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAs 470
1680 TCAGAACAAGAAATCCACCGTTACCATTCACAGCAATTAAGATATTGCTA 1729
1681 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
470 nThrGlnHisThrSerThrIlePheLeuThrGly...LysSerLeuIle 486
1730 CAACCGGCAATACACACAGCTTGATAGC 1758
1731 .....||| |||.....||| |||.....||| |||.....||| |||.....|||
486 hTrpAsnProAsnSerLeuSerValHisSer 495
seq_name: p1r2:S61329

```

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - *Neisseria meningitidis* (frags
C:Species: *Neisseria meningitidis*
A:Variety: NGC16
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61379
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria*
A:Reference number: S61314; MUID:95302961
A:Accession: S61329
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-508 <LOM>
A:Cross-references: EMBL:X82485; NID:g732869; PIDN:CAA57868.1; PID:g732870
C:Superfamily: Iga-specific metalloendopeptidase
*Keywords: hydrolase; metalloproteinase

```
alignment_scores:
  Quality: 870.00
  Ratio: 2.465
  Length: 553
  Gaps: 22
```

```
Percent Similarity: 63.834      Percent Identity: 39.060
alignment_block:
US-09-303-518D-649 x S61329    ..
Align seg 1/1 to: S61329 from: 1 to: 508
```

```
alignment_block:
  nc=00-303 F100
```

Align seg 1/1 to: S61329 from: 1 to: 508

202 GCGAAGATATTGGAGTTTACACAAAAAGGGAGTTGGTCGGCAATC 253
1 AlathraspleusevalLylsasnlsglnl yglinsnllelglyasn1 17
252 AATGACAAAAACCCGATGATTGATTTTCTGTGGTCGGGTAAAC...G 298
17 aleuserasnvalProkmetileaspheserValAlaspralnslnysa 34
299 GCGTGGCGGCGATTGGTGGCGCATATATATGTGGAGCGGTGCACAT... 345
34 rgllealathvalValasproglntlyralValaserValysHlisa 50
346AACGGCGCGATTAAACCGTTGATTT 371
51 LysalaglValAlHsthhrphetyrTtyglcglntlyrnsnlglnhsasnas 67
372 TGGTGGCGAAGAGAAATCCGATCAACATCGTTTACTATATAAATTG 421
67 pvalAlaaslysglnsn...GLTyrAlayalV 78
422 TGAACGAGTAATTTATTAAGACAGGAGCTAAAGCCATCTTATGCGGC 471
78 alglunglnasnslytYglupro.....HlslsyalatpclY 90
472GATTACATATGCGCGTTTGACATA 497
91 AlaserasnleuglYatrgleuglUasptyrAsnmetalarphesnly 107
498 ATTGTGCACAGATGCGAGACCTGTGAATGACCAGTATATGATGGG... 566
107 sphevalthglValAlaProthlealprothraspAlaglyglYglY 124
547 ..CGGAATATATGCATCAAAATATATACCTGACCGGTGTGATATGGG 594
124 eusprhrtYrlyasrplysasnsrpheserSerhevalaayValsgly 140
595 GCAGCGAGGCA.....TATTGGCATGTGATGAAGATGAGCCCAATA 638
141 AlaglyatrglnleuValYtYrglYslgYalAtyrHlsglnlglnlys 157
639 CGCGAAGATTTCAT.....CATTTGGAAGTGGGATTCTTGGC 679
157 n...GlnYsglYtYraspleuayrAspleuserglAlatYratYatY 173
680 TCGTGTGGTGC.....AAT 693
173 lailelaelYthrProtyrYlYasplleasnllleasprlthmetasn 189
694 ACCTTGGCAAAATGATGATCAGGTGGTGGCACAGTCACACTTATGGTATG 743
190 thrglunglYleuileclYrheclYlYasnHlsasntHnHlslYSerlael 206
744 AAAAAATTAAACATAC...CCATATGCGTTTTCACCAACAGAGAGCTCAT 790
206 ucluleuYsglnAlaleuSerGlnasprAlaaleuHrnsnYtYrlyVal 223
791 TTGGGACAGTGGTCACCAATGTTTATCTATGATCCCAAAAGCAAAAG 840
223 eeglyAspserclYserProleuphealAlaPhasrplysglnlYsasngln 239
841 TGTTAATTAATGGGATTTGCAAAACGGGCAACCCCTATATAGGCAAAAG 890
240 tYrValrheleuglYthtYrlyasprYtrprAlaglytYr...Glylys.. 254
891 CAATGGCTTCCACGCTGGTCTAAAGATGGTTCATGATGATCAAAATCTTGG 940

```

255 .....LysSerTrpGlnGluTrpAsnIleTyr. 263
941 CTGGAGATACCCATTCAGTATTCTACGACACGTCATAAATGGAAATAC 990
264 .....LysLys 265
991 TCTTTTACGAC.....GATATATGCGACAGGAAAT 1025
266 GluPheAlaAspLysIleLysGlnArgAspAsnIleGlyThrIleLysG1 282
1026 CAATGCCAATGACGACATTCCTGCTATATAGATATGATATGAAACGAA 1075
282 yaannglyGlnHisIleTrpAsn.....IleThrPheGlyT 294
1076 CCGTCAATGTTTAAATGTTTCTTATCCGAGACACGACAGAACCTGTT 1125
294 hr.....AsnSerHisIleGlySerThrAla..... 302
1126 TATCATCTCGAGTGTGTCACAGTATATGACCCAGCTGATATATG 1175
303 ValArgLeuAlaGlyAsnGluArgAsp.....AlaAsnAsnG1 315
1176 AGAAATATTTCTTATATGACGAGAAAGCGAATTGATCTTATACCA 1225
315 yGlnAsnValThrPhe.....GluAspAsnGlyThrLeuValLeuAspG 330
1226 GCACATCAATCAAGTGTGAGAGATTATTTCCAGAGATTTTACG 1275
330 InaHisIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThr 346
1276 GTCTGGCGTGAATAATACGAA...ACTTGGCAAGCGCGCGCTCTAT 1322
347 ValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspVa 363
1323 CAGTGAACAGTACCGTTACTTGAAGTAAACGCGTGGCAACGACC 1372
363 LAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnLysAsp 380
1373 GCGTGTCCAAATGGCAAGGCGCGCTGCAAGTCAAGCAAGGGGAA 1422
380 rGluAlaLysIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyVal 396
1423 AACCAAGCGTATGACGCGTGGCGAGCGTACAGTCAATTTTGATCAGCA 1472
397 AsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLys 413
1473 GCGACAGCATAAAGGCAAAACAGCCTTAGTGAATCGGCTTGCTCA 1522
413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430
1523 GCGGAGGGGTACGCTGCAACTGAATGCCGATATCAGTTCAACCCGAC 1572
430 eGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 446
1573 AAACCTATTTTCGCTTTCGCGCGAGCGTGTGATTTAAACGGGCATTC 1622
447 AsnLeuTyrPheGlyPheArgGlyLysArgLeuAspAlaAsnGlyAsnAs 463
1623 GCTTTCGTTCCACCGTATTCAAATACGATGAGAGGGCGATGATGCA 1672
463 pLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValA 480
1673 ACCCAATCAAGCAAGAAATCCAGCTTACATTCAGCGCAATTAATAT 1722
480 snHisAsnThrGlyHisIleSerThrIleThrLeuThrGly...LysSer 495
1723 ATTGCTACA 1731
seq_name: p1r2:S61327
496 LeuIleThr 498
seq_documentation_block:

```

```

Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: NCG80
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61327
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314, MIMD:95302961
A:Accession: S61327
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82479; NID:9732883; PIDD:CAA57862.1; PID:9732884
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

alignment_scores:
Quality: 865.00 Length: 534
Ratio: 2.437 Gaps: 20
Percent Similarity: 66.479 Percent Identity: 39.326

alignment_block:
us-09-303-518d-649 x S61327 ..
Align seg 1/1 to: S61327 from: 1 to: 496

202 GCGAAGATATGAGCTTTTACAAACAAAGGCGAGTGTGCGCAATC 251
1 AlaThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17
252 AATGACAAAGCCCGATGATGATTTTCTGCTGCGCTAAC...G 298
17 AlaSerAsnValPrometIleAspPheSerValAlaAspValAlaAsnAlaG 34
299 GCGTGGCGCATTTGGTGGCGCATATATGTCAGCGTGGCGCATAC 348
34 rGThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50
349 GCGGCGTATACACGCTGATTTTGTGCGGAGAGAAATCCGAT... 396
51 LysGlyAspGluIleSerTyrGlyHisIleAsnGlyHisLeuAspVa 67
397 ....CAACATCGTTTACTTATTAATTTGTGAACGCAATTAATTAAG 442
67 lSerAsnAspGluAsnGlnTyrArgSerValAlaGlnAsnAspTyrGluP 84
443 CAGGACATAAGGCCATCTTATGCGGCG.....GATAT 477
84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGlnAspTyr 99
478 CATATGCGCGGTTTGATTAATTTGTACAGATGACGAACCTGTGAAT 527
100 AsnMetAlaArgLeuAsnLysPheValThrGlnValAlaProIleAlaPr 116
528 GACCAGTTATATGATGG...CGGAATATATCATGATCAAAATTAATAC 574
116 oThrSerAlaGlyGlyValGlnValGlyThrGlnPhe..... 144
575 CTAGCCGTGTTGATTTGGGCGAGCGAGCAATATTTGGCATGTGAA 624
133 eGlyPheValArgValGlyAlaGlyThrGlnPhe..... 144
625 GATGAGCCCAATACCGCGAAAGTTCATATCATATTTGCAAGTCGATTC 674
145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160
675 TTGGCTGTTGGTGGCAATACCTTT.....GCACAAATG 709
160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177
710 GATCAGGTGTGGCAGTCACCTTAGTGTGAAAAAATTAACATAGC 759

```

```

177 euAaNGInGluGlyLeuIleGlyPheGlyAspSnsSerLysHSHSSer 193
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
760 CCA.....TATGCTTTTACCAACAGGAG 785
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
194 ProGluLysLeuIysGluValLeuSerGlnAsnAlaLeuThrAspTyrAl 210
786 CTCATTGGGACAGTGCTCACCATTGTTATCTATGATGGCCAAAGC 835
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
210 AValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227
836 AAAAGTGGTTAATTATGGGATTGGCAACGGCAACCCCTATATAGA 885
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
227 ysaTgTrrPValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241
886 AAAAGCATGGCTTCAGCTGGTTCCTAAAGATTGG...TTCATATAGCA 932
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
242 GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLysLys 254
933 AATCTTGGCTGGAGATACCATTCACTATTTACGACACCGCTCAAAATG 982
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
254 sgLurPheAlaAsp.....GluLysGlnGlyTyr. 263
983 GGAATATCTTTTAAAGCATATATATGCGACGAGAAATCAATGCC 1032
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
264 .....AspAsnAlaGlyThrIleLysGlyAsnGly 273
1033 AAACATGAACCAATCTCTGCTAATGATTAATAAACACGACCGTCA 1082
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
274 GlnHSHSTrrPlyThr.....ThrGlyThr..... 282
1083 ATTGTTATGTTTCTTATTCGACACGACGAGAACCTGTTATGATG 1132
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
283 .....AsnSerHisIleGlySerThr.....A 290
1133 CTGACGTGTGTGTACACAGTATTCGACCCAGACTGAATATGAGAAAT 1182
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
290 laValArgLeuAlaAsnAsnGlnLysAsp...AlaAsnAsnGlyGlnAsn 305
1183 ATTTCCTTATTTGACGAGAAAGGCAATGACTTACCAACCAAT 1232
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
306 ValTrpPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsnI 320
1233 CAATCAAGTGCTGGAGATTAATTTCCAAAGAGATTTACGCTCGC 1282
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
320 easnGlnIlyAlaGlyLeuPhePheLysGlyAspTyrThrValLysG 337
1283 CTGAATATACGAA...ACTTGGCAAGCGCGGCTTCATATCAAGCA 1329
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
337 IyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353
1330 GACAGTACCGTACTTGGAAGTAAAGCGCGTGACCAAGCCGCTGTC 1379
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
354 GlyLysLysValValTrrPValValLysAsnProAsnLysAspArgLeuAl 370
1380 CAAATCGCAAGCAAGCGCTGACGCTTCAAGCCAAAGGGGAAACCAAG 1429
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
370 alysiIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGln 387
1430 GCTGATTCAGCGTGGCGAGCGTACAGCTATTTGATCAGCAGCAGAC 1479
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
387 IyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403
1480 GATAAAGCAAAAACAAGCCTTAGAGAAATGGCTGGTCAGCGGCAAG 1529
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyTyr 420
1530 GGTACGCTCACTGAATGCGATATATCACTTCAACCCCGCAAACTGT 1579
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
420 gArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuT 437
1580 ATTTCGCTTCGCGCGGCGAGCTTGATTAACGCGCATTCGCTTGG 1629
    ||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

437 yrrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453
1630 TTCACCGATTTTCAAAATACCGTGAAGGGGAGGATGATTTGCAACCAAA 1679
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHis 470
1680 TCAGACCAAAAGAAATCCACCGCTTACCATTAACGCAATTAAGATATGCTA 1729
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
470 nrrGlnHisAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486
1730 CA 1731
486 hr 486
seq_name: p1r2:S61320
seq_documentation_block:
  IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
  C:Species: Neisseria meningitidis
  A:Variety: NGC65
  C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
  C:Accession: S61320
  R:Rimboldi, H.; Poulsen, K.; Mogens, K.
  M:1. Microbiol. 15, 495-506, 1995
  A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
  A:Reference number: S61314; MIMD:95302961
  A:Accession: S61320
  A>Status: preliminary; nucleic acid sequence not shown
  A:Molecule type: DNA
  A:Residues: 1-508 <LOM>
  A:Cross-references: EMBL:X82484; NID:g732871; PDB:CA57867.1; PID:g732872
  C:Superfamily: IGA-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloproteinase

alignment_scores:
  Quality: 865.00 Length: 558
  Ratio: 2.403 Gaps: 22
  Percent Similarity: 64.516 Percent Identity: 38.889

alignment_block:
  US-09-303-518D-649 x S61320 ..
  Align seg 1/1 to: S61320 from: 1 to: 508

202 GCGAAGATATTGAGGTTTACACAAAAGGAGGAGTGGCGCAATC 251
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  1 AlatrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGCAAAAGCCCCGATGATGATTTTCTGCGTGTGCGCTAAC...G 298
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
  17 aleuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGGCAATTTGGTGGCGCATATATTTGTAGCGTGGCAGAT... 345
    : :||| :||| :||| :||| :||| :||| :||| :||| :|||
  34 rGleAlaIatThrValValAspProGlnTyrAlaValSerValLysAla 50
346 .....AACGGGCGCTATTAACAACTTATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlnHisAsn 67
372 TGGTGGCGAAGGAAGAAATCCGATCAACATCGTTTACTTAAATATG 421
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  67 pValAlaAspLysGlnAsn.....GluTyrArgValVal 78
422 TGAACGGAATATTATTA...CGAGGAGCTAAA 453
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  78 alGluGlnAsnAsnTyrGluTrpHisLysAlaTrpSerLaserAsnLeu 94
454 GGCATCTTATGGCGGCGATTTATGATGCGCGCTTGCATTAATTTGT 503
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
  95 GlyArgLeu.....GluAspTyrAsnMetAlaArgPheAsnLysPheVa 109

```



```

299 GCGTGGCGGCGATGTGGGCGCATATATGTGAGCGGCGACATAC 348
    :::::::::::::::::::: ||||| ||||| |||
34 rghrleuthrvalileasprglttvalalaservallyshsl 50
    :::::::::::::::::::: ||||| ||||| |||
349 GCGGCGCTATACAAAGCTGATTTGGTGGGGAAGAAGAAATCCGAT 396
    :::::::::::::::::::: ||||| ||||| |||
51 lsgslasprcileserlytlyglnshslasnglnshleuaprla 67
    :::::::::::::::::::: ||||| ||||| |||
397 .....CAACATCGTTTACTATTAATTTGCAAGCGAATATATTAAAG 442
    :::::::::::::::::::: ||||| ||||| |||
67 lserasnasprgluasnglttvalrgersevalalaclnasnsprlyt 84
    :::::::::::::::::::: ||||| ||||| |||
443 CAGGGACTAAAGGCGCATTCCTTAGGCGGC.....GATTAAT 477
    :::::::::::::::::::: ||||| ||||| |||
84 roslasnasntprhls...HsGLYAsnGlnGlyArgLeuGlnsPrlyt 99
    :::::::::::::::::::: ||||| ||||| |||
478 CATATCCCGCGTTTGCATTAATTTGTCACAGATGCAAGAACTGTGAAT 527
    :::::::::::::::::::: ||||| ||||| |||
100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116
    :::::::::::::::::::: ||||| ||||| |||
528 GACCACTTATATGATGGG...CGGAATATATATCATCAAAATATTAC 574
    :::::::::::::::::::: ||||| ||||| |||
116 othrserlagaLygLyglValGluThrlytLysAspLysAsnArgPheS 133
    :::::::::::::::::::: ||||| ||||| |||
575 CTGACCGGTGTGATTTGGGCGGACGACCAATATTGGCATCTGATGAA 624
    :::::::::::::::::::: ||||| ||||| |||
133 etcluphevalargvalglvalaglythrGlnPhe..... 144
    :::::::::::::::::::: ||||| ||||| |||
625 GATGACCCCATACCGGGAAGTTCATATCATATATGCAAGTGGTATTC 674
    :::::::::::::::::::: ||||| ||||| |||
145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160
    :::::::::::::::::::: ||||| ||||| |||
675 TTGGCTGTGGTGGCAATACCTTT.....GCACAAATG 709
    :::::::::::::::::::: ||||| ||||| |||
160 gtyrAlaIleAlaGlyThrProTyrGlnAspValAlaSnAlaThrSerSnL 177
    :::::::::::::::::::: ||||| ||||| |||
710 GATCGAGTGTGGCACAGTCAACTAGTAAATTAATTAACATAGC 759
    :::::::::::::::::::: ||||| ||||| |||
177 euasnglnGlnGlyLeuIleGlyPheGlyAspAsnSerLyshslSer 193
    :::::::::::::::::::: ||||| ||||| |||
760 CCA.....TAGGTTTTTACCAACAGAGG 785
    :::::::::::::::::::: ||||| ||||| |||
194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210
    :::::::::::::::::::: ||||| ||||| |||
786 CTCATTTGGCGACAGTGGCTCCACAAAGTTATCTATGATGCCCAAGC 835
    :::::::::::::::::::: ||||| ||||| |||
210 avallleuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGlu 227
    :::::::::::::::::::: ||||| ||||| |||
836 AAAAGTGTAAATTAATGGGTATTGCAACGGGCAACCCCTATATAGGA 885
    :::::::::::::::::::: ||||| ||||| |||
227 ysargttrvalrPheleuGlyAlaTyrAspTyrtrPalaGlyTyr..... 241
    :::::::::::::::::::: ||||| ||||| |||
886 AAAAGCAATGGCTTCACAGTGTGCTGAAGATTTG...TTCTATGATGA 932
    :::::::::::::::::::: ||||| ||||| |||
242 GlnLysAsnSerTrpGln.....GluTyrAsnIleTyrLysLys 254
    :::::::::::::::::::: ||||| ||||| |||
933 AATCTTCTGCAAGTACCAATTCAAGTATCTACGAACACAGCTCAAAATG 982
    :::::::::::::::::::: ||||| ||||| |||
254 sglurPheAlaAsp.....GluIleLysGlnArg. 263
    :::::::::::::::::::: ||||| ||||| |||
983 GGAATACTCTTTTAACGAGATAATTAAGCACAGGAAAAATCAATGCC 1032
    :::::::::::::::::::: ||||| ||||| |||
264 .....AspsnAlaGlyThrIleLysGlyTyrGly 273
    :::::::::::::::::::: ||||| ||||| |||
1033 AAACATGACAAATCTCTGCTAATAGATTAATAAACAGACAGACCTTCA 1082
    :::::::::::::::::::: ||||| ||||| |||
274 GluHisThrTrpLysThr.....ThrGlyThr..... 282
    :::::::::::::::::::: ||||| ||||| |||
1083 ATTGTTAATGTCTTTATCCGAGACAGAAAGAAACCTGTATTATCATG 1132
    :::::::::::::::::::: ||||| ||||| |||
283 .....AsnSerHisIleGlySerThrAlaValArgLeuAlaGlyAsng 297
    :::::::::::::::::::: ||||| ||||| |||

```

```

1133 CTCAGGTGGTGTCAACAGTTATGACCCAGACGTAATTAATGAGAAAT 1182
    :::::::::::::::::::: ||||| ||||| |||
297 luhrglyAla.....AsnAsnGlyGlnsn 305
    :::::::::::::::::::: ||||| ||||| |||
1183 ATTTCCGTTATTGACGAGAAAGCGCAATGATACTTACCAGCAACAT 1232
    :::::::::::::::::::: ||||| ||||| |||
306 ValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsnI 320
    :::::::::::::::::::: ||||| ||||| |||
1233 CAATCAAGGTGTGAGAGATTAATTTCCAGAGAAATTTTACGGTCCG 1282
    :::::::::::::::::::: ||||| ||||| |||
320 eaSnglnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337
    :::::::::::::::::::: ||||| ||||| |||
1283 CTGAATTAACGAA...ACTTGCAAGCGCGCGGTTCATATGACGTGAA 1329
    :::::::::::::::::::: ||||| ||||| |||
337 lvalAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353
    :::::::::::::::::::: ||||| ||||| |||
1330 GACAGTACCGTTACTTGAAGTAAGCGCGTGGGCAAGACCGCTGTG 1379
    :::::::::::::::::::: ||||| ||||| |||
354 GlyLysLysValValThrPheValLysAsnProAsnGlyAspArgLeuAl 370
    :::::::::::::::::::: ||||| ||||| |||
1380 CAAATGCGCAAGGACGCGTCCAGCTTCAAGCCAAAGGGAACCAAG 1429
    :::::::::::::::::::: ||||| ||||| |||
370 alYsIleGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGln 387
    :::::::::::::::::::: ||||| ||||| |||
1430 GCTCGATCAGCGTGGGCGAGGTACAGTCATTTTGATCAGACGAC 1479
    :::::::::::::::::::: ||||| ||||| |||
387 lYglLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403
    :::::::::::::::::::: ||||| ||||| |||
1480 GATTAAGGCAAAACCAAGCCTTTAGTGAATTCGCTGTGTCAGCGCAG 1529
    :::::::::::::::::::: ||||| ||||| |||
404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValIserGlyAr 420
    :::::::::::::::::::: ||||| ||||| |||
1530 GGTACGCGTGCACAGTGAATGCCGATATCACTTCAACCCGACAACTCT 1579
    :::::::::::::::::::: ||||| ||||| |||
420 garGlyrLeuValIleuAsnSerProAspGlnIleAsnProAsnLeuT 437
    :::::::::::::::::::: ||||| ||||| |||
1580 ATTTGCGCTTTCGCGGAGGTTTGATTAACGGGCAATTCGCTTCG 1629
    :::::::::::::::::::: ||||| ||||| |||
437 ytrPheGlyrPheArgGlyLysLysLeuAspAlaAsnGlyAsnAspLeuThr 453
    :::::::::::::::::::: ||||| ||||| |||
1630 TTCCACCGTATTCAAATATCCGATGAAAGGGCGATGATTCACACCA 1679
    :::::::::::::::::::: ||||| ||||| |||
454 PheGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAs 470
    :::::::::::::::::::: ||||| ||||| |||
1680 TCAAGCAAAAGATCCACCGTTACCATTTACGCAATTAAGATTTGCTA 1729
    :::::::::::::::::::: ||||| ||||| |||
470 nThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486
    :::::::::::::::::::: ||||| ||||| |||
1730 CAACCGCAATTAACACAGCTTG 1752
    :::::::::::::::::::: ||||| ||||| |||
486 hrAlaProGlnAsnLeuSerVal 493
    :::::::::::::::::::: ||||| ||||| |||

```

seq_name: p1r2:S61330

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (f

C:Species: Neisseria meningitidis

A:Variety: SM1166

C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S61330

R:Domholt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neis

A:Reference number: S61314; MUID:95502961

A:Accession: S61330

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-508 <LDM>

A:Cross-references: EMBL:X82486; NID:q732860; PIDN:CAAS7869.1; PID:q732861

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

alignment_scores: Quality: 856.00 Length: 551
 Ratio: 2.404 Gaps: 22
 Percent Similarity: 64.610 Percent Identity: 38.657

alignment_block:
 us-09-303-518d-649 x S61330 ..

Align seg 1/1 to: S61330 from: 1 to: 508

```

202 GCGAAGATTTTGAAGTTTACAAACAAAAGGGAGTTGGTCGCAATC 251
    ||| |||.....||| |||.....|||.....: |||.....:
1  AlathrSplenserValLysasnLysGlnGlnsnlleGlysnal 17
252 AATGACAAAAGCCCGATGATTTTCTGCTGCTGCGTAAAC...G 298
    :|||.....: ||| |||.....: ||| |||.....: |||
17 aleuSerAsnValProMetIleAspSerValAlaAspValAsnLysa 34
299 GCGTGGCGCATTTGGTGGCGATCATATATTTGAGCGTGCACAT... 345
    :|||.....: ||| |||.....: ||| |||.....: |||
34 rgleAlaThrValValAspProGlnValAlaSerValLysHisAla 50
346 .....: ||| |||.....: ||| |||.....: |||
51 LysAlaGlnValHisThrPheTyrTyrGlnTyrAsnGlnHisAsnAs 67
372 TGTGCGGAAGAGAAATCCCGATCAACATCGTTTACTATTAATTTG 421
    |||.....: ||| |||.....: ||| |||.....: |||
67 pValAlaAspLysGlnAsn.....GlnTyrTyrValVal 78
422 TGAACCGAATATTTATTA.....GACGAGCATTA 453
    |||.....: ||| |||.....: ||| |||.....: |||
78 aIlglnGlnAsnTyrGlnProHisLysAlaTyrPserAlaSerAsnLeu 94
454 GGCATTCCTTATGCGGCGATTCATATGCGCGTTGCATTAATTTGT 503
    |||.....: ||| |||.....: ||| |||.....: |||
95 GlnTyrGln.....GlnAspTyrAsnMetAlaArgPheAsnLysPheVa 109
504 CACAGATCCAGACCTGTTGAATGACCACTTATATGATGGG...CGGA 550
    |||.....: ||| |||.....: ||| |||.....: |||
109 lThrGlnValAlaProIleAlaProThrAspAlaGlnLysLysLeuAspT 126
551 AATATGATCATCAAAATATATACCTGACCGTTCGATGGGCGGAGCG 600
    |||.....: ||| |||.....: ||| |||.....: |||
126 hTyrLysAspLysAsnArgPheSerSerPheValArgValAlaGln 142
601 AGGCAA.....TATTGGCATCTGATGAGATGACCCCAATTAACCGGA 644
    |||.....: ||| |||.....: ||| |||.....: |||
143 ArgGlnLeuValTyrGlnLysGlnAlaTyrHisGlnGlnLysAsn...G 158
645 AAGTTCATAT.....CATATGCAAGTGGCATTCCTTGGCGCTGG 685
    |||.....: ||| |||.....: ||| |||.....: |||
158 nLysGlnTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIle 175
686 GTGGC.....AATACCTTT 699
    |||.....: ||| |||.....: ||| |||.....: |||
175 lalGlnThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThrGln 191
700 GCACAAATATGATCAGGTGGTGGCACGATCACTTATGAGTAAAT 749
    :||| |||.....: ||| |||.....: ||| |||.....: |||
192 GlnLeuIleGlnPheGlnLysAsnHisAsnThrHisTyrSerAlaGln 208
750 TAAACATATC...CCATATGTTTATTACCAACAGAGGCGATTTGGCG 796
    :|||.....: ||| |||.....: ||| |||.....: |||
208 nLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlnValLeuGln 225
797 ACAGTGGCTCACCAATGTTTATATGATCCCAAAAGCAAAAGTGTTA 846
    ||| |||.....: ||| |||.....: ||| |||.....: |||
225 spSerGlnLysPheLeuPheAlaPheAspLysGlnLysAsnGlnTyrVal 241
847 ATTATGGGCTATGCAACGCGCAACCCCTATATAGAAAAAGCAAGG 896
    :||| |||.....: ||| |||.....: ||| |||.....: |||
242 PheLeuGlnTyrThrTyrAspTyrTrrPalaGlnTyr...GlnLys..... 254
  
```

```

897 CTTCAGCTGGTTCGTAAGATGGTTCATATGATGAATCTTCTGAG 946
    |||.....: ||| |||.....: ||| |||.....: |||
255 .....: ||| |||.....: ||| |||.....: |||
947 ATACCATTCAGTATTCACAGACACGCTCAAAATGGCAATACTCTTT 996
    ||| |||.....: ||| |||.....: ||| |||.....: |||
264 .....: ||| |||.....: ||| |||.....: |||
997 AACGAC.....GATATATATGCGACAGCAAAATCAATGC 1031
    ||| |||.....: ||| |||.....: ||| |||.....: |||
268 AlaAspLysIleLysGlnArgAspAsnAlaGlnTyrIleLysGlnAla 284
1032 CAACATGACACCAATCTCTGCTTATATGATTAACCAACGACGTTTC 1081
    :||| |||.....: ||| |||.....: ||| |||.....: |||
284 yGlnHisThrPasn.....IleThrPheGlnThr... 294
1082 AATGTTTAAATGTTCTTATCCGAGACAGACAGACAACTGTTATCAT 1131
    ||| |||.....: ||| |||.....: ||| |||.....: |||
295 .....: ||| |||.....: ||| |||.....: |||
1132 GCTGACAGTGGTTCACACGATATGACCCAGACTGAATATGAGAA 1181
    ||| |||.....: ||| |||.....: ||| |||.....: |||
305 LeuAlaGlnLysGlnLysArgAsp.....AlaAsnAsnGlnGlnAs 317
1182 TATTCCTTATTTGACGAAGAAAGCGAATTTGATTTACCAACA 1231
    ||| |||.....: ||| |||.....: ||| |||.....: |||
317 nValThrPhe.....GlnAspAsnGlnTyrLeuValLeuAspGlnAsnI 332
1232 TCAATCAAGTCTGAGAGATTTATTTCCAGAGAGATTTACGCTTCG 1281
    ||| |||.....: ||| |||.....: ||| |||.....: |||
332 lAsnGlnGlnLysGlnLysLeuPhePheLysGlnLysPyrThrValLys 348
1282 CCTGAAATAAGAA...ACTTGGCAAGCGCGCGCTTCATATCATGTA 1328
    ||| |||.....: ||| |||.....: ||| |||.....: |||
349 GlnAlaAsnAsnAspIleThrThrLeuGlnLysAlaGlnLysPheVal 365
1329 AGACGTACCGTTCCTTCTGAAAGTAACGCGCTGGCAACGACCCCTGT 1378
    :||| |||.....: ||| |||.....: ||| |||.....: |||
365 pGlnLysLysValValTrrPglValLysAsnProAsnGlnLysPheVal 382
1379 CCAAAATCGGCAAGGACGCTGACGCTGACGCAACGCAAGGAAACCA 1428
    ||| |||.....: ||| |||.....: ||| |||.....: |||
382 lAsnIleGlnLysGlnTyrLeuGlnLysGlnLysGlnLysGlnLysGln 398
1429 GGTGATGACGCTGGCGGACGCTGACGCTGATTTGGATCAGGAGCA 1478
    ||| |||.....: ||| |||.....: ||| |||.....: |||
399 GlnGlnLeuLysValLysGlnTyrValIleLeuAsnGlnLysAlaAs 415
1479 CGATTAAGCAAAACAAACGCTTTAGTGAATCGGCTTGGTCAGGCGCA 1528
    :||| |||.....: ||| |||.....: ||| |||.....: |||
415 pSerAsnGlnLysValSerAlaPheSerGlnValGlnLysIleValArgGln 432
1529 GGGTACCGGCGCAACGTAATCCGATATCATGTTCAACCCGCAAACTC 1578
    ||| |||.....: ||| |||.....: ||| |||.....: |||
432 rGlnThrLeuValLeuAsnSerProAspGlnIleLeuAsnProAsnLeu 448
1579 TATTCGCTTTCGCGGAGCTTTGATTAACGGCATTCGCTTTC 1628
    ||| |||.....: ||| |||.....: ||| |||.....: |||
449 TyrPheGlnLysPheArgGlnLysArgLeuAspAlaAsnGlnLysAsn 465
1629 GTTTCACCGCTATTCAAATACGATGAAAGGGGAGGATTTGCAACCA 1678
    :||| |||.....: ||| |||.....: ||| |||.....: |||
465 rPheGlnHisIleArgAsnValAspGlnGlnLysAlaArgIleValAsnHis 482
1679 ATCAAGCAAGAATCCACCGTATACCATTCAGCAGCAATATGATTTGCT 1728
    ||| |||.....: ||| |||.....: ||| |||.....: |||
482 snThrAspArgAlaSerThrIleThrLeuThrGln...LysSerLeuIle 497
1729 ACA 1731
    |||
498 Thr 498
  
```


1541 AACTGAATGCCGATATCAATTCACACCCGACAACTCTATTTCGGCTTT 1590
 |||||.....
 435 alleuansnsersersenglnlleasnProAspnleuylrphelgylphe 451
 1591 CGCGGGCGAGCTTGGATTAAACGGGCATTCGCTTCCTCCACCGTAT 1640
 |||||.....
 452 Arg91ygllyArpleuaspalaasnlglyasnAspleuThrPhegluhstl 468
 1641 TCMAAATACCGATGAAGGGCGATGTGTCAACACCAATCAAGCAAAAG 1690
 |||||.....
 468 eArgAsnValAspLuglValAlaArglleValAsnHlsAsnThrGlyHlsA 485
 1691 AATCCACCGTTACCATTCACGCAATPAACATATTCCTCA 1731
 |||||.....
 485 laserThrleuThrleuThrGly...LysSerleuThrleuThr 497

seq_name: p1r2:S61332

seq_documentation_block:
 Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strain
 N/Altenate names: Igal protease
 C:Species: Haemophilus influenzae
 A:Variety: HK635
 C>Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
 C/Accession: S61332
 R:Lomholt, H.; Poulsen, K.; Mogens, K.
 Mol. Microbiol. 15, 495-506, 1995
 A:Title: Comparative characterization of the iga gene encoding Igal protease in Neisseria
 A:Reference number: S61314; MUIDID:95302961
 A:Accession: S61332
 A>Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-462 <LOM>
 A:Cross-references: EMBL:X82488; NID:g732714; PIDN:CA57871.1; PID:g732715
 A:Experimental source: strain HK635
 A:Note: the authors did not translate the codon for residue 462
 C:Gene: iga
 C:Superfamily: Iga-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 712.50 Length: 506
 Ratio: 2.213 Gaps: 21
 Percent Similarity: 63.636 Percent Identity: 36.166

alignment_block:
 US-09-303-518D-649 x S61332 ..

Align seg 1/1 to: S61332 from: 1 to: 462

202 GCGAAGATATTGAGTTTACAAACAAAGGGAGTGTGCGCAATC 251
 |||.....
 1 AlaThrAsnValGluValAlaArgAspLysAsnHlsSerleuGlyAsnAl 17
 252 AATGACAAAGCC...CCGATGATTGATTTTCTGTGCTGCGGTAAC. 297
 |||||.....
 17 AleuProAsnGlyIlePheMetIleAspPheSerValValAspValAsnL 34
 298 ..GGCGTGGCGGCGATGTGTGCGGCAATATATATGTGAGCGTGACAT 345
 |||||.....
 34 ysaArglleGlyThrleuValAspProGlnTyrlleValSerValLysHls 50
 346 AACGGCGGCTATPAACAACCTGATTTT.....GTCG 377
 |||||.....
 51 AlaHlsGlnIleTyMetAsn...AspPheTyrlrPheGlyHlsTyAsnGlyHi 66
 378 GGAAGAAGAAATCCGATCAACATCGTTTACTTATAAATGTGAAC 427
 |||||.....
 66 sArgAspValSerAspAspGluAsnLys.....TySerValValThrng 81

428 GGAATATATTATAAGCAGGAGCTAAAGCCAT.....CCTTATGGCGC 471
 |||||.....
 81 lnsAsnValAsnProAsnGluAsnTrpHlsValAspLysArgLeuAsp 97
 472 GATTATCATATGCGCGCTTGCATTAATTTGTCACAGATGCAAGAACCTG 521
 |||||.....
 98 AspTyAsnMetProArGleuAsnLysPheValThrGluValAlaProTh 114
 522 TCAAAAGACACAGTTATATGATGGG...CGGAATATATTCGCAAAATA 568
 |||||.....
 114 rThrProThrleuAlaGlyAspAspLeuGluThrTyrllysAspLysGlu 131
 569 ATTACCCCTGACCGCTTTCGTATTTGGGCGAGGACCA.....TATTGG 612
 |||||.....
 131 ystTyrlleuSerPheValArgValGlyAlaGlyArgGlnleuValTyrlu 147
 613 CGATTCGATGA.....GATGAGCCCAATPAACCGCAAGTTCATATCA 656
 |||||.....
 148 LysGlySerHlsHlsValGluAspLysGluHlsGlyGluAspLeuLysAs 164
 657 TATTGACAGTGGCTATTCCTTGCTGCTGCTGTCGCAATACCTTTGCA... 702
 |||||.....
 164 pleuSerAlaIleTyArgTyrlAlleGlyThrProTyrlLysGlyI 181
 703CAAAATGATCAGGTGGTGGCACAGCTCACTTAGTAGT 741
 |||||.....
 181 leAsnIleAspProSerGlnSerLysGlyLeuIleGlyPheGlyAsp 197
 742 GAAAAATTTAAACAT..... 756
 |||||.....
 198 SerArgGluAspHlsValIleAsnSerLysThrleuLeuSerGlnAspPr 214
 757 ...AGCCATATGCTTTTACCACACGAGGCTCATTTGGCGACAGTG 802
 |||||.....
 214 oleuThrAsnTyrlGlyValleu.....GlyAspSerG 225
 803 GCTACCAAGTCTTATCTATGATGACCCCAAGCAAAAGTGTAAATTAAT 852
 |||||.....
 225 LysSerProLeuPheAlaPheAspLysGlnlmsnLysTyrlValPheIle 241
 853 GGGTATTTGCAACGGGCACCCCTAT.....ATAGAA 887
 |||||.....
 242 Gly.....ProTyrlThrTyrlPheAlaGlyTyrlGly 252
 888 AAGCATGCTTCACGCTGCTCTAAAGATGG...TCTATGATGA 934
 |||||.....
 252 LysSerTrp.....GlnGluTrpAsnIleTyrlLysSerG 264
 935 TCTTGTGAGATACCCATCACTATTCAGAACCCAGTCAAAATGGG 984
 |||||.....
 264 lnsPheThrLysAspVal..... 269
 985 AATATCTCTTTAAACGACGATATATATGCGACAGCAAAATCAATGCAA 1034
 |||||.....
 270LeuAsnLysAspSerAlaGlyLeuLeuLysGlyAsnTrgl 283
 1035 ACATGAACACATTCCTGCGCTAAT...AGATTAACAAACGAAACCTTC 1081
 |||||.....
 283 nTyrlAsnTrpHlsSerAsnGlyAsnThrSerMetIleSerAsnLysSerG 300
 1082 AATGTTTAATGTTTCTTATTCGAGACGACGCAAGAAACCTGTTATCAT 1131
 |||||.....
 300 lLeuLeuGluValAsnLeuPheAspAsnSerLys..... 311
 1132 GCTGACAGTGGTGTCAACAGTATTCAGCCACAGCTCAATATGAGAAA 1181
 |||||.....
 312HisThrAsnArgGluLysAlaAsnTyrlGlyLysSe 323
 1182 TATTCTTTATTTGACGAAGAAAGCGAATGTATCTTACCGACACA 1231
 |||||.....
 323 rValThrPhe.....GlnGlyAsnGlyThrleuThrleuLysAsnSerI 338
 1232 TCATCAAGGTGCTGAGAGATTATATTCACAAAGATTTTACCGTC... 1278

```

|||||
338 leasngllylaglyglyleuphehgluglyasntrythrvalglu 354
1279 TCGGCTGAATAATACGAAAGCTGGCAAGGCGCGCTTCATTCAGTCA 1328
355 glyserasrapsnilevaltrpasnlglylaglyleservalalagl 371
1329 AGACAGTACCGTTACTGTGAAGTAACAGCGCTGGCAAGCAGCCCTGT 1378
371 uglyllysthrvalthrtrpysvalhnsasnproglinsersapargleua 388
1379 CCAAAATCGGCAAGGCAAGCGCTGACCTTCACCTTACCAAGGCAACCAA 1428
388 lalylstleglyllysglythrleuilevalgluglyllysglyllyasnlys 404
1429 GCGTCATCAGCGTGGCGAGCGGTACAGTCATTTTGATCAGCAGCCACA 1478
405 glyserleuylsvalglyaspolythrvalilleleulysglnglnlaas 421
1479 CGATAAAGCAAAAACACAGCCTTTAGTGAATCGGCTTGTCAGCGCA 1528
421 palasnasnlyllyvallysallapheserclnvalglyilevalserglya 438
1529 GGGGTACGCGTCACTGAATGCCGATATCAGTTCACCCCGCAAACTC 1578
438 rgsertthrvalvalleuasnaspplysglvalaspasnsertile 454
1579 TATTTCGGCTTTCGCGCG 1596
455 tyrpheglyphearggly 460
seq_name: p1r2:S61333

seq_documentation_block:
19A specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strai
N:Alternate names: 19A1 protease
C:Species: Haemophilus influenzae
A:Variety: HK284
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61333
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the 19a gene encoding 19A1 protease in Neisseri
A:Reference number: S61314; MUID:95302961
A:Accession: S61333
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-471 <LOW>
A:Cross-references: EMBL:X82487; NID:9773205; PIDN:CAA57870.1; PID:9773206
A:Experimental source: strain HK284
C:Genetics:
A:Gene: 19a
C:Superfamily: 19A-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 677.50 Length: 520
Ratio: 2.111 Gaps: 24
Percent Similarity: 61.731 Percent Identity: 35.000

alignment_block:
US-09-303-518D-649 x S61333 ..
Align seg 1/1 to: S61333 from: 1 to: 471

202 CGGAAGATTATTGAGTTTACACAAAAAGGAGTGTGCGCAATC 251
||| .....||| ||| .....|||
1 Alathrasnvalgluvalargaspplysasnnglnserleuylseral 17
252 AATGACAAAGCC...CCGATGATGATTTTCTGTCGTGTCGCGTAAC 297
||| .....||| ||| .....|||
17 aleuproasnlglyleuprometileaspheservalvalaspvalasn 34

```

```

298 ..GGCGTGGCGGCAATGTTGGGCGATCATATATTGTGACCGTGCAAT 345
||| .....||| ||| .....|||
34 ysatrgileglythrleuvalaspproclntlyrillevalserllyshs 50
346 AAC.....GGCGCTATACAAAGCTTGA 368
51 Alathlshshsilleasnaspheptyrphleglyhnslyasn..... 64
369 TTTTGTGCGGCAAGAAAGAAATCCGATCAACATCGTTTACTATTAA 418
65 ..Glyhnsargaspvalseraspaspplasnlys.....tyrser 78
419 TTGTGAACGGAATATATTAAAGCAGGACTAAAGCCAT.....CCT 462
78 alvalthrclnasnasnvallyspasnglnasprrphtlshvalasp 94
463 TATGGCGGATATATCATATGCGCGTTTGCAATTAATTGTGACAGATGC 512
||| .....||| ||| .....|||
95 Argleuasnaspptyrasnmetproatgleuasnlysphevalthrgluva 111
513 AGAAGCTGTGAATAGACCGATTATGATGAGG...CGAATATATATCG 559
111 lalaprothrthrprothrleuvalaglyaspaspplasnlythrlysa 128
560 ATGAAATATATTCCTGACCGCTGTTCGATTTGGGCGAGCGCAATAT 609
||| .....||| ||| .....|||
128 splysgllystlyrproserphevalargvalglalaglythrclnph 144
610 TGCGATCTGAT.....GAAATGAGCCCAATTAACCGCA 644
145 valtyrclnlysglyserlyrvalglulysrthrtrntrargasnasn 161
645 AAGTTATATCATATGCAAGTGCCTATTCCTGCTGCTGTCGACATA 694
||| .....||| ||| .....|||
161 pleyllysphe...leuaspplualatyrargtyrallecllyglythr 177
695 CTTTGCA.....CAAAATGATCAGGTGGTGGCACACTC 729
177 rotyrclnlyllyleasnilleaspserglnserlyllyleuile 193
730 AACTTAGTAGTAGAAATTAATTAACAT..... 756
||| .....||| ||| .....|||
194 glypheglyaspserarglunshisvalilleaspalalysrthrleu 210
757 .....AGCCATATGTTTTCACCAAGAGGCTCAT 790
210 uSerlinspproleuthrasntryrlyvalleu..... 221
791 TTGGGAGAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAG 840
||| .....||| ||| .....|||
222 ..glyaspserglyserproleuphealaphaspplysglnglnasnlys 237
841 TGCTTAATTAATGGGCTATGCAAGCGGCAACCCCTAT..... 879
||| .....||| ||| .....|||
238 tyrvalphelegly.....protyrthrtyrtpal 248
880 ....ATAGAAAGCAAT.....GGCTTCAGCTGCTCGTAAGAT 919
||| .....||| ||| .....|||
248 aglytyrcllyllysllysertrpnlgnlutrpnasnlleryrlyllyasp 264
920 GGTTCATGATGAATCTTTGCTGAGATACCCATTCAGTATTCAGCA 969
||| .....||| ||| .....|||
265 ..phealaspasnillellyllyargaspasnlagluvalaval..... 278
970 CCAGCTGAATGGGAATATCTTTTAACGACAGATATATATGCGCAGC 1019
||| .....||| ||| .....|||
279 Proheserthrserglutryhsttrpthrasnthrshasnshln 295
1020 AAAAATC...AATGCCAATGATGACCAAT.....TCTCTGC 1054
||| .....||| ||| .....|||
295 rgulilellyasnthrshpshthrillerthrvalthrleuproserasp 312

```

```

1055 CTAATGATTTAAAAACGACGACGTTCAATGTTAATGTTCTTATGCC 1104
      |||||
312 rosnargyleu..... 315
1105 GAGACAGCAGAGACCTGTTATTCATGCTGCAGGTGCTCAC...AG 1151
      |||||
316 .....Valasnhedl 319
1152 TTATGACCCAGACTGATATATGAGAAAATATTTCTTATTTATGAC.... 1197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
319 nglnlysglnhlsleuglnthrclglnasnvalthrphespspsert 336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1198 .....GAGAGAAAAGCGGAATTGACTTACACAGCAATCAATCAAGGT 1242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 hrasnansnglysglythrleuileuaspshisileasnngly 352
1243 GCTGAGAGATTATTTCCAAAGAGATTTCAGCGTCGCGCAAAATATA 1292
      |||||
353 Alaglylyleuphepnepelysglyasntrglnvallysglysrhras 369
1293 CGAA...ACTTGGCAGCGCGCGCGCTTCATATCATGTAAGACATACCG 1339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
369 pspilletthrtrpvalglyglyileaspvalalaglulglylsqlyuv 386
1340 TTAATTGAAAGTAAAGCGCGTGGCAACGCCCTGTCGCAAAATGCGC 1389
      |||||
386 alvaltrpysvalhisasnproglulysasphisleualalysllely 402
1390 AAAGGCGCGTGCAGCTTCAAGCCAAAGGGGAAACCAAGCGCTCGATCAG 1439
      |||||
403 Lysglythrleuilevalgluglylysglyasnslsglyserleu 419
1440 CGTGGGCGAGCGTACAGTCAATTTGGATCAGCAGCAGCAGCATAAAGCA 1489
      |||||
419 svalglyspglythrvalvalleuylsglnglnthrasmelyserly 435
1490 AAAAACAAGCCCTTAGTAATCGCTGTGTCAGCGCAGCGGATACGGTG 1539
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 ..gluhisalapheklaservalglyilevalserglyargserthrval 451
1540 CAACCTGAATGCCGATATCACTTCAACCCGACAACTCTATTTGCGCTT 1589
      |||||
452 Valleunasnaspplysglnvalasprianserlietyrpneglyph 468
1590 TCGCGCGCGCA 1599
      |||||
468 earglygly 471

seq_name: plr2:T00317

seq_documentation_block:
probable serine proteinase esp, extracellular - Escherichia coli plasmid p0157
N/Alternate names: putative exoprotein-precursor
C/Species: Escherichia coli
C/Date: 01-Feb-1999 #sequence-revision 01-Feb-1999 #text-change 08-Dec-2000
C/Accession: T00317; T42120
R/Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yatsudo, H.C., Kubota, S., Shinagawa, H.
DNA Res. 5, 1-9, 1998
A/Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7, strain EHEC O157:H7, substrain RMD 0509952
A/Reference number: 214127; MUID:98290540
A/Accession: T00317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1300 <MAX>
A/Cross-references: EMBL:AB011549; NID:94589740; PIDN:BA31826.1; PID:93337077
A/Experimental source: strain EHEC O157:H7, substrain RMD 0509952
R/Burland, V., Shao, Y., Perna, N.T., Plunkett, G., Sofia, H.J., Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A/Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7, strain EHEC O157:H7, substrain RMD 0509952
A/Reference number: 222068; MUID:98391744
A/Accession: T42120
A/Status: preliminary; translated from GB/EMBL/DBJ

```

```

A/Molecule type: DNA
A/Residues: 1-1300 <BUP>
A/Cross-references: EMBL:AF074613; PIDN:AACT00088.1
A/Experimental source: strain EDL933; serotype O157:H7
A/Genetics:
A/Genes: esp
A/Genome: plasmid p0157
C/Superfamily: 1g4-specific metalloendopeptidase
C/Keywords: serine proteinase

```

alignment_scores:

```

Quality: 635.00 Length: 1506
Ratio: 0.863 Gaps: 67
Percent Similarity: 48.871 Percent Identity: 22.112

```

alignment_block:

```

us-09-303-518d-649 x T00317 ..

```

```

Align seg 1/1 to: T00317 from: 1 to: 1300

```

```

82 TTAGCCATATGCTGTCGTCGATTCCTCCCAAGCGTGGCGGAC 131
      |||||
40 LeuAlaLeuCysphe...leuylleuenglnserlyrserpheal 55
132 CACTTATTTCCGATCACTACCAATAC.....TATGCCACTTTG 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 aserglnmetasplleserasnphetylrlleargaspyrmetaspheal 72
173 CGAAATATAAGCAAGTTTGGACGCGGCGGCAAAATGATGAGTTTAC 222
      |||||
72 laglnasnlysglylephenglnalaglyalathasnleuileuileval 88
223 AACCAAAAAGGCGAGTTGGTCGCGCAATCAATGACAAAAGCCCGATGAT 272
      |||||
89 LysLysAspLysSerThrleu.....LysleuProgluValIleProPhepr 103
273 TGAATTTTCTGTGTCGCCGTAACGCGTGGCGGATGTTGGCGATC 322
      |||||
103 onspheiserprovalalalasnlysglyserthrtrserllleglyglya 120
323 AATATTTTGGAGCGTGGCAGATAC...GGCGGCTATACAGCTTGAT 369
      |||||
120 latyrserliethralathrhlsasnthrlylasnshislerValala 136
370 TTTGTCGCGGAGAGAAATCCGATCAACATCGTTTACTTATTAAT 419
      |||||
137 ThrGlnasotrprglyanser.....Thryrlysgl 147
420 TGTGAACGCAATATTTATAACAGAGACTAAGGCCATCTTATGCGG 469
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
147 nrhrasprtrpsnthr.....SerHispro..... 155
470 GCGATTTCATATGCGCGCTTTCATTAATTTGTC..... 504
156 ..AspHealavalaserAthrleuaspshysphevalvalglutrrargly 171
505 ...ACAGATCAGAACCTGTTGAATGACGAGTTATATGATGGGGGAA 551
      |||||
172 Alathrluglylaasplileserleuserlysglnlelaleuuguar 188
552 ATATATGATCAATAAATTAATACCTGACCGTGTTCGATTTGGGCGAGCA 601
      |||||
188 gtyr.....GlyValasntrlyrlysglylulysleuilealpheal 203
602 GGCATATTTGCGATCGATGAGATGAGCCCAATAACGC..... 642
      |||||
203 rgalaglyserclgylvalvalserValylsasnnglyargliethrpro 219
643 .....GAAGTCATAT.....CA 656
      |||||
220 pheasnngluvalserlyrlysglulmetleuasnnglyserphevalhi 236

```

657 TATTGCAAGTGGCTATCTTGGCTCGTT...GGTGGCAATACCTTTGCAC 703
 236 sllleasprprSerIyLrLeuIleuThrAsnAsnGlnrPheaspG 253
 704 AAATGGATCAGGTGGTGACAGCTCAACTAGTAGTGAATAATTTAA 753
 253 Lurheasn.....AsnIleAsnSerGln..... 260
 754 CATAGCCCATATGTTTTTACACAGAGAGGCTCATTTGGCGACAGTGG 803
 261GlyAspSerI 264
 804 CTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGGTAAATTAATG 853
 264 ySerIaleuPheValIyAsnAsnGlnIyLysIyLysrPrValIalA 281
 854 GGGATTTCAAAACGGGCAACCCCTATATAGCAAAACAAATGGCTTCAG 903
 281 LyThrVal..... 283
 904 CTGGTTCTAAAGATGGTCTATGTAATCTTTCGTGAGATACCA 953
 284TPrGlyIleYAsnYrAlaAsnGlyLysAsnH 295
 954 TTCAGTATCTAGCAACCGCTCA.....AATGGGA 985
 295 sAlaIalAySerIySerIyStrAsnGlnThrIleAspAsnLeuLysAsnL 312
 986 AATACCTTTTAAACGAGATATATATGACACAGCAAAATCATGCCAA 1035
 312 ySTySerIyAsnValAspMetSerClyAla..... 322
 1036 CATACACAAATTCCTCGCCTAATAGATTAATAACACAGACCGTTCAATT 1085
 323GlnValAlaThrIleGlu.. 328
 1086 GTTAAATGTTCTTATCCGAGACAGACAGAACCTTTATCATGCTG 1135
 329 ...AsnGlyLysLeuThrIleGlyThrGlySerAsp..... 338
 1136 CAGGTGGTGTCAACAGTTATCGACCCAGACTGATATATGAGAAATATT 1185
 339ThrThrAspIleLysAsnLys 345
 1186 TCCTTTATGAGAGAAAGCCAAATGATCTTACCAGCAACATCAA 1235
 346 AspLeuIlePheThrIyLylAspIleLeuLysSerSerPheAs 362
 1236 TCAAGGTGCTGAGATTTATTTCCAAAGAGATTTTACGGCTCGCGCTG 1285
 362 rAsnGlyAlaGlyLysLeuValrPheAsnAspLysThrIyArGyAla 379
 1286 AAATATACGAA...ACTTGGCAAGCGCGGCGCTCATATCACTGAAGAC 1332
 379 snIyLysAspAspPheThrPheLysGlyAlaGlyValAspThrArGAsnGly 395
 1333 AGTACCGTTACTTGAAGATTAAGCGCGTGGCAACACACCGCTGTCCAA 1382
 396 SerThrValGluTrpAsnIleTrgYrAspAsnLysAspAsnLeuHsY 412
 1383 AATGGCAAAAGCGACGCTGCACGTTTCAAGCCAAAGGGGAAACCAAGGCG 1432
 412 sIleGlyAspGlyThrLeuAspValArG...LysThrGlnAsnThr...A 427
 1433 CGATCAGCGTGGCGACGCTACAGCTATTTGGATACAGCAGGACAGAT 1482
 427 snIeulYstrIyGlnIyLysLeuValIleLeu..... 437
 1483 AAAGGCAAAAGCAAGCCTTAGTGAATGGCGTTGGTCAGCGGCAAGG 1532
 438 ...GlyAlaGlyLysThrPheAsnAsnIleYrIleThrSerIyAspI 453
 1533 TACGGTGCACACTGAATGCCGATTAATCAGTTCAACCCCGACAA..... 1575

453 yThrValrLeuAsnAlaGlyAsnAlaLeuSerGlyCylGlyYrAsn 470
 1576 ...CTCATTTGGCGTTTCGGCGAGCGATTTGGATTTAAAGGGCATTCG 1623
 470 yLlePhePheAlaLysAsnGlyYrIleLeuAspLeuAsnIyYrAsn 486
 1624 CTTTGGTTCCACCGTATTAATACCGATGAGAGGGCGCATGTGTCAA 1673
 487 GlnSerPheAsnLysIleAlaIalThrAspSerGlyAlaValIleThrAs 503
 1674 CCAACATCAACAAAGATTCACACCGTTACCATTTACAGCAATTAAGATA 1723
 503 n...ThrSerThrLysSerIleLeuSerLeuAsnAsnThrIalAspTr 519
 1724 TTGCTACACCGGACATTAACACAGCTTGATAGCAAAAGAAATTCGCC 1773
 519 yTleYrYrHsGlyAsn..... 524
 1774 TACACCGTTGGTTTGGCCGAAAGATACGACCAAAACGAACGGCGCT 1823
 525IleAsnGlyAsnLe 529
 1824 CAACCTTGT.....TACAGCGCGCGCGAGAACAGCGCACCTGCTGC 1867
 529 uAspValLeuGlnHsIleGlnIyLysLysGlnAsnArGArIleIlel 546
 1868 TTTCGGCGGAGACAAATTTAAACGGCAACATCACGCAACAAACGGCAAA 1917
 546 euAspGlyGlyValAspThrThrAsnAspIleSerLeuArGAsnThrGln 562
 1918 CTGTTTTTCAGCGGCGACCAACACCGGACGCTTACAATCATTTAAACGA 1967
 563 LeuSerIleGlnIyAlaIalThrGlnIleAlaIle...YrArGAspG 578
 1968 CCAATGGTGCAGAAAGAGAGCGCATTCCT..... 1995
 578 yAlaPheSerCysSerLeuProIalrPheMetArGpHeuIySerIySera 595
 1996CGCGGGGAATCGTGGGACAC 2019
 595 sPrYrValAlaGlyMetGlnAsnThrGlnAlaAspAlaValLysGlnAsn 611
 2020GA 2021
 612 GlyAsnAlaTyLysThrAsnAsnAlaValSerAspLeuSerGlnProAs 628
 2022 CTGATCAACCGACATTTAAAGCGGAAATTCGCAATTAAGGGCGAC 2071
 628 rTrGlnThrGlnIyThrPheArGPrHeGlyThrLeuHsIyLeuGlnAsnSera 645
 2072 AGCGGCGTTCGCCGCAATGTGTCGCAAAAGTGAAGCGCATTTGG 2121
 645 sPheSerValIleYrAsn...AlaAsnValIleGlyAspIleGlnAla 660
 2122 AGCAATCAACGCGCAAGCAGTTTGGTGGTGGACCGGATCAAAAGCCACAC 2171
 661 SerLys..... 662
 2172 AATCTGTACAGTTTCGACTGACGCGTGCACAAATGTGTGCAAAAAA 2221
 662 662
 2222 CCATTAACGAGATTAAGATGTCCTTCATGACTGAACGACCATCAGC 2271
 663SerAsnIleThrIleGlyAspThrThr 671
 2272 GGCATTTGCATCTTGCAGATCAGGCTCATTTAAATCAGACAGG..... 2316
 672 AlaTyIleAspLeu.....HisAlaGlyLysAsnIleThrGlyAspI 686
 2317CTTGCACACATCAACGGCAATCTTGAATGGCGC 2353
 2353


```

3905 TTTCAGACGC..... 3915
1234 euAlaasnGLygluThrValleuArAspAlaSerIylgluIysArgIle 1250
3916 ATGGAGGCAAAATCCGCCGCCGCTGTCATTACGCATTACAGCAGC 3965
1251 LysglYgluIysAspSerArGmetLeuMetSerValIgluAsnAlaGI 1267
3966 ATACCGC.....GCCGCTTCG 3982
1267 uileAlgAspAsnValArPheglYleuGluphegluIysSerAlaPheG 1284
3983 GCGGATTGCGCATCGAACCGCACATCGCGCAACGCGATTTCGTCAA 4032
1284 LysTyrAsnValAspAsnAlaVal..... 1292
4033 AAGCGGATTACCGCTAC 4050
1293 AsnAlaAsnPhaArgTyr 1298

```

seq_name: p1r2:B41500

seq_documentation_block:

19A-specific metalloendopeptidase (EC 3.4.24.13) type 2 - Haemophilus influenzae (fragme
N:Alternate names: Immunoglobulin A1 proteinase type 2
C:Species: Haemophilus influenzae
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Dec-2000
R:Grundy, F.J.; Plant, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A:Title: Localization of the cleavage site specificity determinant of Haemophilus influe
A:Reference number: A41500; M01D:90129281
A:Accession: B41500
A:Status: not compared with conceptual translation
A:Residues: 1-384 <GRU>
C:Comment: This 19A1 proteinase is classified as type 2 because it cleaves at a proline
C:Superfamily: 19A-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 400.00 Length: 426
Ratio: 1.633 Gaps: 20
Percent Similarity: 57.512 Percent Identity: 29.108

alignment_block:

US-09-303-518D-649 x B41500 ..

Align seg 1/1 to: B41500 from: 1 to: 384

```

64 GCGTTCGCGCTGCTTACTTAGCCATATGCCGTGCTGGCATTCCTCC 113
5 LysPheIysLeuAsnPhelIeAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCCCGCGGCGGACACACTTATTCGGCATCACTACCAATACATC 163
21 oTyrThrcIuAlaIaLeuValArGAspArpValAspTyrGlnIlePheA 38
164 GCGACTTGGCGAAATAAAGCAAGTTTCAGTCGGGCGGAAAGATAT 213
38 rGAspPheAlaGluAsnLysglYArPheSerValIgluAlaThrAsnVal 54
214 GAGGTTCACACAAAAAGGAGGAGTTGTCGGCAATCAATGACAAAGC 263
55 GluValArGAspLysAsnAsnHisSerLeuGlYAsnValleuProAsnGI 71
264 C...CCGATGATTGATTTTCTGTGTGTCGCTAAC...GGCGTGGCG 307
71 yIlePheMetIleAspPheSerValValAspValAsnLysArgIleGIYT 88
308 CATTCGTCGGCGATCATATATTTGTGAGCGTCGACATAAGCGCGCTAT 357

```

```

88 hrLeuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104
358 AACACGTTGATTT.....GTCGCGAAGAGAA 389
105 MetAsn...AspPheTyrPheglYhIStyrAsnGlnHisArgAspValSe 120
390 TCCGATCAACATCGTTTACTTATTAATGTGAACGAAATATATTA 439
120 rAspArgLysAsnLys.....TyrSerValValThrGlnAsnAsnValA 135
440 AACACGAGACTAAAGCCAT.....CCTTATGCGCGCATATATCATG 483
135 snProAsnGlnLysnTrpHisValAspLysArgLeuAspAspTyrAsnMet 151
484 CCGCTTTCATTAATTTTCACAGATGAGAACCTGTGAATACACAG 533
152 ProArgLeuAsnLysPheValThrGluValAlaProThrThrProThrLe 168
534 TTATATGATGGG...CGGAAATATATCATCAATCAATATACCTGACC 580
168 uAlaGluAspAspLeuGluThrTyrLysAspLysGluLysTyrLeuSerP 185
581 GTTTCGTTGCGGCGAGCGAGCA...TATGCGCATTCATGAA 624
185 heValArGValGluAlaGluArgGlnLeuValTyrGluLysGlnSerHis 201
625 .....GATGAGCCCAATACCCGCAAGTTCAATATCATATGCAAGTC 668
202 HisValGluAspLysGlnHisGluLysLysPheLysAspLeuSerAlaI 218
669 GATTCCTTGGCGCTGGTGGCAATACCTTGCA.....C 703
218 aTyrTrpTyrAlaIleGlyGlyThrProTyrLysGlyIleAsnIleAsp 235
704 AAATGATACAGTGTGGACACAGTCACTAGTAGGAAATTAATAA 753
235 roSerGlnSerLysLysGlyLeuIleGlyAspSerArgGluAsp 251
754 CMT.....AGCCCAT 764
252 HisValIleAsnSerLysThrLeuLeuSerGlnAspProLeuThrAsnTy 268
765 TGGTTTTCACCAACAGAGGCTCATTTGGCGACAGTGGTCCCAATC 814
268 rGlyValLeu.....GlyAspSerGlySerProLeuP 279
815 TTATCTATGATGCCCAAAAGCAAGTGTAAATTAATGGGATATGCA 864
279 heAlaPheAspLysGlnGlnAsnLysTrpPheIleGly..... 292
865 ACGGCAACCCCTAT.....ATAGGAAACCAATGCTT 899
293 .....ProTyrTrpTyrTTPAlaGlyTyrGlyLysLysSer..... 304
900 CCAGCTGGTTGTAAAGATGG...TTCATGATGAATAATTTGGTCGAG 946
305 .....ThrGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 318
947 ATACCCATTCAGTATTCTACGACCGCTCAAAATGGAAATACCTCTTT 996
318 spVal.....Leu 320
997 AACGACATATAATATGACACAGAAAAATCAATGCCAAATGAAACAA 1046
321 AsnLysAspSerAlaGlyLeuLeuLysGlnHisThrGlnTyrAsnTrpH 337
1047 TTCCTCCCTAAT...AGATTAACAAACAGAACCGTTCAATTGTTAAG 1093
337 rSerAsnGluYAsnThrSerMetIleSerAsnGlnSerGlnLeuLeuGly 354
1094 TTCTTATATCCGACAGACAGAACCGTTTATCATCTCTCAGAGTGT 1143
354 aLAsnLeuPheAspAsnSerLys..... 361

```

```

1144 GTCACAGCTATTCGACCCGACTGAATATGAGAGAAATATTCCTTT 1193
      ::::: ||| :: ||| ::::: |||
362 ...HISTHtAsnArgGluysAlaAsnArgGlyLysSecValThrPhe.. 376
      ::::: ||| :: ||| ::::: |||
1194 TGACGCAAGAGAAAAGCGAATGTACT 1221
      ::::: ||| ||| |||
377 ...GlnGlyAsnGlyThrLeuThrLeu 384

seq_name: pir2:C91068

```

seq_documentation_block:
 hypothetical protein ECS3515 [imported] - Escherichia coli (strain O157:H7, substrain RHD)
 C|Species: Escherichia coli
 C|Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C|Accession: C91068
 R|Hayashi, T., Makino, K., Onishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G.
 gasaara, N., Yasunaga, T., Kuwara, S., Shiba, T., Hattori, M., Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A|Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
 A|Reference number: A99629; MUID:21156231; PMID:11258796
 A|Accession: C91068
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1-1571 <HAY>
 A|Cross-references: GB|BA000007; PIDN:BA836938.1; PID:q13362986; GSPDB:GN00154
 A|Experimental source: strain O157:H7, substrain RHD 0509952
 C|Genetics:
 A|Gene: ECS3515

```
alignment_scores:      Length: 1689
      Quality: 367.00
      Ratio: 0.470      Gaps: 88
Percent Similarity: 46.240 Percent Identity: 20.130
```

Align seg 1/1 to: C91068 from: 1 to: 1571

```

268 ATGATGATTTTCTGCTGCTGGTAAACCGGCGGCAATGATGGG 317
   : : : : : : : : : : : : : : : : : : : : : : : : :
60 LeuIleSerAlaLeuIleTyrAlaProProGlyMetAlaAlaPheThr 76
   : : : : : : : : : : : : : : : : : : : : : : : : :
318 CGATCAATATATTTGAGCGGTGGCAATAAC.....GGCGGTATA 358
   ||| : : : : : : : : : : : : : : : : : : : : : : :
76 oAsp.....ValIleGlyValAlaAspGluThrValAspGlySerG 91
   : : : : : : : : : : : : : : : : : : : : : : : : :
359 ACAACGTGATTTTGGTGGAGGAAGAAACCCGATCAATCGTTT 408
   : : : : : : : : : : : : : : : : : : : : : : : : :
91 IaaIValAsp.....GluArgGlyThrThrAsnAsnThrHis..... 103
   : : : : : : : : : : : : : : : : : : : : : : : : :
409 ACTTATAAATTTGTAAACGGAATATTATAAGCAGGACTAAAGCCA 458
   ||| : : : : : : : : : : : : : : : : : : : : : : :
104 .....IleIleAsnHisGlyGlnGlnAsnVal..... 112
   : : : : : : : : : : : : : : : : : : : : : : : : :
459 TCCTTATGGCGCGCATATATCATATGGCGGTTTGCAATAATTGTGCAG 508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 ...TyrGlyGly.....ValSerA 118
   : : : : : : : : : : : : : : : : : : : : : : : : :
509 ATGCAGAACCTTTGAAATGACCGAGTTATATGAT...GGCGCG..... 549
   : : : : : : : : : : : : : : : : : : : : : : : : :
118 snGlySerLeuIleGlnSerGlyGlyTyrClnAspValGlyArgHisAsn 134
   : : : : : : : : : : : : : : : : : : : : : : : : :
550 AAATATATGCATCAAAATATTAACCTGACCGCGTTGCGTATGGCGCAG 599
   : : : : : : : : : : : : : : : : : : : : : : : : :
135 AsnTyrValGlyGlnSerAsn.....AsnThrThrIleAsnGly 148
   : : : : : : : : : : : : : : : : : : : : : : : : :
600 CAGGCAATATTTGGCGATTCGATGAAGAT..... 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 yArgGln.....SerIleHisAspGlyGlyIleSerThrGlyThrI 167
   : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

628 .....GAGCCCAATTAACCGGAAAGTTCATPACATATGTCAGTCGGTAT 111 672
162 leileugserglyasnglnaspyalTyllyselglylleseransgly 178
673 TCTTGCGTCGGTGGCCCAATTAACCTTTGGCAAAATGATCAAGTGGTCG 722
179 ThrThrIlelysglyalaseratgyalIgulgylseralasnol 195
723 CACA...GTCACTTAGTGGTGAATAAATTAACATACCAATAGTAT 769
195 yThrleuleaspolyglyserglnilevalTysvalinglnghnsal 211
770 TTTTACCAACGAGGGGTCTATTGGCAGACGTGGTCACCAATGTTTATC 819
212 .....AspglyThrThrIleasnlysserglyserglnaspyalVal 225
820 TATGATGCCCAAAAGCAAAAGTGTATTATTAAGGGGTATTCGAAACGGG 869
226 GlnglyserleualathrasnthrThrIleasnly.....ol 238
870 CAACCCCAATATAGA.....AAAGCA 892
238 yArgGlnTyrAlaGlnGlnserThrValGluThrThrIlelysaang 255
893 ATGGCTTCACCTGGTTCGTAAGATTTGGTCTTATGATGAATCTTGCT 942
255 TylGlnGlnArgAlaTylTyrGlnserArgAlaleuasprThrThrIleGlu 271
943 GGAGATACCACCTGACATATTCAGTAACCAACGTCMAATGCGAA..... 987
272 GlyGlnTyrGlnserleu.....AsnerlysserTh 282
987 ..... 987
282 rAlaLysasnThrGlnIleTyrserGlyGlyThrGlnIleleleaspsant 299
988 .....TACCTTTTAAAC.....GACGAT 1005
299 hNserSeraspValIleGluValTyrserGlyGlyValleuaspa 315
1006 AATTAATGCGACAGGAAATGACATGCCAAACATGACAGCAATTCCTGCC 1055
316 SerGlyGlyThrAlaThrAsnValThrGlnHisaspGlyAlaIleuLy 332
1056 TAATAGATTAAAAACACAGACCGTTCATTTGTTAAATGTTCTTATCCG 1105
332 sTrhAsnthrasnglyThrThrValSerGlyThrAsn.....Serg 346
1106 AGACACACAGAAACCTGTTTATCATGCTGCA..... 1137
346 lnclylAarheserIenHisasnHisValAlaaspaasnValleuLeuglu 362
1138 ...GGTGGT.....GTCAACAGTAT..... 1155
363 AsnGlyGlnHisleuasprlleasnaIatryGlySerAlaasnlyThrIl 379
1156 .....C 1156
379 eilelyasprlysglyThmetSerValleuthrasnaIatysalaaspa 396
1157 GACCCAGACTGAATATAGA..... 1176
396 lathrAarGlleaspaasnGlyGlyValMetaspyAlaIaglyAsnaIathr 412
1177 .....GAAATATTTCCTTATGACGAGG 1202
413 AsnthrIleleasnGlyGlyThrGlnasnIleasasnTylGlyIleal 429
1203 AAAAGCGAATTGATCTTACAGCAACAATCAATCAAGGT..... 1242
429 aThrGly.....ThrasnleasnserGlyThrGlnasnI 441
1243 .....GCTGAGCATATATATTCGAAAGGAGATTTTACGGTCTCGCGCTGA 1287

```


441 LelysSerGlyGly.....LysAlaAspThrThrIleIleSerSer 454
1288 AATTAAC.....GAAACTTGGCAAGCGCGCGCTTCAT 1322
455 GlySerArgGlnValValGluLysAspGlyThrAlaIleGlySerAsn 471
1323 CAGTGAAGACAGTACCGCTTACTTGGAAAGTAAACGGCGGTGCAAGACGAC 1371
471 eSerAlaGlyGlySerIleuIleValTyrThrGlyGlyIleAlaHisGly 488
1372CGCTGTCCAAATCGCAAGACGACG 1398
488 AlaSngInGluThrGlySerAlaLeuValAlaAsnThrGlyAlaGlyThr 504
1399CTGCACGTTCAAGCCAAAGG..... 1419
505 AspIleGluGlyTyrAsnLysLeuSerHisPheThrIleThrGlyGly 521
1420GAAACCAAGCGCTGATCAGCGGTGGCGACG 1450
521 uAlaAsnTyrValValLeuGluAsnThrGlyGluLeuThrValAlaAla 538
1451 GTACAGTCATTTGGATCAGCAGCAGATAAAGCAA..... 1491
538 ysthSerAlaLysAsnThrThrIleAspAlaGlyLysLeuIleVal 554
1492AAACAAACCTTTAGTGAATCGGCTTGTGCAGCGCAGGGGTACGCT 1538
555 GluLysGluAlaLysThrAspSerThrArgLeuAsnAsnGlyLysAla 571
1539 GCACAGCAATGCCGAT.....AATCAGTTCACACCGCG 1570
571 uGluValGlnAspGlyGlyLysAlaLysHisValGluGlnInSerGly 588
1571 ACAACTCTATTTCCGCTTCCGGCGGACGTTG.....GATTAAAC 1614
588 LysAlaLeuIleAlaSerThrThrSerGlyThrLeuIleGluGlyThrAsn 604
1615 GGGCATTCGCTTTCGTTCCACCGTATTCAAATACCGATGAAGGGCGCAT 1664
605 SerTyrGlyAspAlaPheTyr...IleArgAsnSerGluAlaLysAsnVal 620
1665 GATTGTCAACCAACAAATCAGACAAAGATCCACCGTTACCATTCAGGCA 1714
620 ValLeu.....GluAsnAlaGlySerLeuThrValValThrGly 634
1715 ATAAAGATATTGCTACACCGGCAATTAACACACGTTGATGCAAAAAA 1764
634 eArgAlaValAspThrIleIleAsnAlaAsn.....GlyLysPhe 647
1765 GAAATTCCTACACAGGTTGGTTGGCGGAAAGAT.....ACGACCA 1808
648 AspVal.....TyrGly...LysAspValGlyThrValLe 658
1809 AACGAACGGCGGCTCAACCTTGTTCACAGCCGCCGACAGACGCCCA 1858
658 uAsnSerAlaGlyThrGlnThrIleThrIleAlaSerAlaThrSerAspLys 675
1859 CCCTGCTCTCTTCGCGGCAACAAATTAACGGCAGCATCAGCAACA 1908
675 LaAsnIleLysGlyLysGlnThrValTyrGlyLeuAlaThrGluAla 691
1909 AAC.....GCCAACTGTTTTCAGCGGCGACACCAACGCCGAC 1949
692 AsnIleGluSerGlyGlnIleValAspGlySerThrGluLys.. 707
1950 CTACAATCATTTAAACGAC..... 1968
708 ...ThrHisLeuAsnGlyLysGlnThrValGlnAsnTyrGlyLysA 723
1969CATTTGGTGCAGAAAAAGAGCGC 1989

723 IAlaLeuAsnThrAspIleValSerGlyLeuGlnIleMetAlaAsnGly 739
1990 ATTCTCGCGGGGAAATCTGTG...TGGCAGACACGTCGATCAACCGCAG 2036
740 ThrAlaGluGlySerIleLeuAsnGlyLysSerGlnIleValAsnGlu 756
2037 ATTTAAACGGGAAACTTCCAAATTAAGCGGACAGCGGTGTTTCC 2086
756 yGlyLeuAlaGluAsnSerValLeuAsnAspGlyGlyThrLeuAspVal 773
2087 GCATGTTTCCAAAGCGGCAAGTCGATTTGGCATTCAGCATCCGCCCA 2136
773 rGluLysGlySerAlaThrGly.....IleGlnInSerSerGln 786
2137 GCAGTTTGTGTGTCGACCGCGATCAAGCCACACAAATCTGACGTTG 2186
787 GlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArgAl 803
2187 GGACTGACGCGGTGACAAATTTGTCGAAAAA..... 2220
803 asp.....GlyValAlaPheSerIleGluGlnGlyAlaAlaAsnAsn 818
2221ACCATTCACGACGAT..... 2235
818 IeLeuLeuAlaAsnGlyGlyValLeuThrValGluSerAspThrSer 834
2236AAAGTATTCG 2246
835 AspLysThrGlnValAsnThrGlyGlyArgGluIleValLysThrLysAl 851
2247 TTCATTGACATAAGACCGCATCAGCGGCAATGTCATGTT..... 2286
851 eThrAlaThrGlyThrThrLeuThrGlyGluGlnIleValAlaGluLys 868
2287 ..GCCGATCAGCGCTCATTTAATCTCACAGCGCTTGCACACCTCAAGCGC 2334
868 AlaAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnThr..... 881
2335 AATCTTACTGCAAAATGGCGATACA...CGTTATACAGTCAACCCAAACGC 2381
882 ...ValSerAlaAsnGlyGluAlaIleLysThrThrIleAsnGluGly 897
2382 CACC.....CAAAAGCGCAACCTTAGC...CTGTGGCGCATG 2416
897 YThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValAlaAsn 914
2417 CCCAAGCACATTTAATCAAGCCACATTAACGGCAACACATCGGCTTCG 2466
914 eArgAlaAlaLeuGlnThrSerThrAlaAsnGly...IleGluIleSer 929
2467 GGCAT.....GCTCATTTAATCTAAGCGACACCGCCGTACAAA 2507
930 GlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAs 946
2508 CGGCACTGTGACGCTTCCGGCAACGCTAAGGCAACGTAAGCATTCG 2557
946 nMetLeuLeuGluAsnGlyLysAsnLeuValIleAlaGlyThrGluAla 963
2558 CACTCAACGGTATGTCCTCCCTAGCCGATTAAGCAGATTTCCATTTGAA 2607
963 IaArgAspSerThrValGlyLysGly...AlaMetGlnAsnGlnGly 978
2608 AGCAGCGCTTTACCGGCAAAATCAGCGGCGGCGAGATACGCAATTACA 2657
979 GlnAspSerAlaThrLysValAsnSerGlyGlnTyrThrLeuGlyArg 995
2658 CTTAAAGACAGCAATGACGCTGCGTCAGCAGCAATTA..... 2700
995 gSerLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnVala 1012
2700 2700
1012 IaGlyGlyThrAlaIleValTyrAlaGlyThrLeuAlaAspAlaSerVal 1028

2701GCAATTTAACTT.....GACAGCCAC 2726
 1029 SerGIYAlaThrGIYSerLeuSerLeuMetThrProArgAspAsnValTh 1045
 2727 C...ATTACCTCAATTCCGGCTATGCG...CAGATGGCGGCGGGCGG 2770
 1045 rProValLysLeuGIYAlaIleArgIleThrAspSerAlaThrLeuT 1062
 2771 AAACCGCAGTGGCAGACATGCGCGCGCGCGCTTGGCGCGTGGCGG 2820
 1062 hrIleGIYAsnGIYValAspThrThrLeuAlaAspLeuThrAlaAlaSer 1078
 2821 CGT.....TCCTATTA..... 2823
 1079 ArgGIYSerValThrLeuAsnSerAsnAsnSerCysAlaGIYThrSerAs 1095
 2824TCCTATTA..... 2832
 1095 ncYsGIYThrArgValAsnSerLeuLeuAsnAspGIYAsnValIYrL 1112
 2833TCGTTACACCGCCCACTTGGTAGATCCGTTTCAACACG 2874
 1112 euSerAlaGIYThrAlaAlaProAlaThrThrAsnGIYIleYrAsnThr 1128
 2875 CTGACGGTAAACGGCAATGAAAGGTGAGGAACATTCCGCTTTATGTC 2924
 1129 LeuThrThrAsn...GIYLeuSerGIYSerGIYAsnPhetYrLeuHisThr 1144
 2925 GGAACCTCTCGCTACCGCACGACGCAAAATGAACTGGCGGAAAGTTCCG 2974
 1144 rAsnValAlaGIYSerArgGIYAspGIYLeuValAlaAsnAsnAlaTr 1161
 2975 AAGCACTTACACCTTGGCGGTCAACATACCGGCAACGAACTGGCAGC 3024
 1161 hrGIYAsnPhetYrIlePhetYrAlaIleAspThrGIYValSerProGIYSer 1177
 3025 CTCGAACATTTGACGCTAGTGTGAGGAAGAAAGCAAAACCGCTGTCCGA 3074
 1178 AspAspAlaMetThrLeuVal.....LysThrGIYGIYGI 1189
 3075 AAACCTTAATTTACCTGCAAAACGAA.....CAGCTCATCCGCGG 3118
 1189 yAspAlaSerPhetSerLeuGIYAsnThrGIYPhetValAspLeuGIY 1206
 3119 CGTGGCGTTACCACTCATCCGCAAGACGCG.....GAGTTCCGCGTG 3162
 1206 hrTYrGIYIYr...ValLeuYsSerAspGIYAsnSerAsnThrAsnLeu 1221
 3163 CATTAATCCGTCGTAAGAAACAAGAGCTTTCGCAAACTGGGCAAGCAGA 3212
 1222 ThrAsnAspValYsPro.AsnProAspProAsnProAsnProAsnProA 1238
 3213 ACCCAAAAACAGCGGCAAAAAGACACAGCCGCAAAAGCTTACCGCGCTGA 3262
 1238 snProLys..... 1240
 3263 TTGGCGCGGCGGCGATGCGCTGCAAAAGACAGAAAGCTTGGCGAACC 3312
 1241ProAsp.. 1242
 3313 GCCCGGACAGCAGGCGGGAATGTGCGATTATGACAGCGGAGGAAGA 3362
 1242 1242
 3363 GAAAAAACGGTGCAGCGGATTAAGACACCGCTTGGGAAACAGCGCG 3412
 1242 1242
 3413 AAGCGGAACCGGCGGCTACACAGCGCTTCCCGCGCGCGCGCGCGC 3462
 1243ProLysProAsp.ProLysProAspProL 1252

3463 CGCGGATTTGCCCAACTGCAACCCCAACCGGACCCCAACCGCAGCG 3512
 1252 yPro.Asp...ProThrProGIYProThrProThrProValProGIY 1267
 3513 CGACCTGATCAGCGGTTATGCCAATAGCGGTTTGAGNGAATTTCCGCCA 3562
 1267 sarGIYThrPro.....SerThrAlaAlaValLeuAsnMetAlaIat 1282
 3563 CGCTCAACAGCGTTTTC.....GCCGTACAGAGCAATTA 3597
 1282 hrLeuProLeuValPhetAspAlaGIYLeuAsnSerIleArgGIYArgLeu 1298
 3598 GACCGCGTATTTGCCGGAAGACCGCGCAGCGCGCTTGGACAGCGCAT 3647
 1299 AsnIleMetLysAlaSerProHisAsnAsnValTrp...GIYAlaThr 1314
 3648 CCGGACACCAACACTACCGTTCCGAGATTTCCGCGCTACCGCCAC 3697
 1314 rTYrAsnThrArgAsnAsnValThrThrAspAlaGIYAlaGIYPhetGIY 1331
 3698 AAACGACCTGCGCCAAATCGTATGCAAAAAACCTC..... 3735
 1331 InThr.....LeuThrGIYMetThrValGIYIleAspSerProAsn 1344
 3736GCGACGCGCGCGCTCGCGCATCTGTTTGGCAACACG 3773
 1345 AspIleProGIYIleAlaThrLeuGIYAlaPhetMetGIYTrSerHI 1361
 3774 GACCGAAACACCTTCGACGACGCG..... 3798
 1361 serHisIleGIYPhetAspArgGIYGIYHisGIYSerValGIYSerYrL 1378
 3799 ..ATCGCAACTGCGCAGCGCTTGGCCACGCGCGCTTTC..... 3837
 1378 erLeuGIYGIYrAlaSerTrpGIYHisGIYSerGIYPhetYrLeuAsp 1394
 3838 GGGCAATACGCGATCGACAGCTTC.....TCATCGCATCGACGCG 3878
 1395 GIYValValLysLeuAsnArgPhetGIYSerAsnValAlaGIYLysMetSe 1411
 3879 GGGCGCGGTTTTCAGCGCGGACGCTT...TCAGCGCATCGAGCA 3925
 1411 rSerGIYGIYAlaAlaAsnGIYSerYrHisSerAsnGIYLeuGIY 1428
 3926 AAATCCGCGCGCGCTGTCATTCAGGATTCAGGACAGATACCGCGCG 3975
 1428 Is.....IleGIYThrGIYMetArgPhe 1435
 3976 GGTTCGGGCGGATTCGCAACCGCACATCGCGGCAACCGCTATTT 4025
 1436 ThrAspGIYAsnThrAsnLeuThrProTYrAlaSerLeuThrGIYPhetH 1452
 4026 CGTCAAAAACGCGATTACCGC.....TAGAAAACGTCATA 4063
 1452 rAlaAspAsnProGIYTrHisLeuSerAsnGIYMetGIYSerYr 1469
 4064 TCGCCACACCGCGCTTGCATTCACCGCTACCGC...GGGCGCATTAAG 4110
 1469 alaSpThrArgSerIle.....TYrArgLeuLeuGIYAlaThr 1481
 4111 GCAGATTATTCATTCAACCGCGGCAACACTTTCCATCCAGCTTATTT 4160
 1482 LeuSerYrAsnMetArgLeuGIYAsnGIYMetGIYIleGIYProTrpPhe 1498
 4161 GAGCGTGTCTATACGAGTCCGCTTCGGCAAAAGTCCGACACGCGTCA 4210
 1498 uLYsAlaIleAlaValArgYsGIYPhetValAspAspAsnArgValYsVala 1515
 4211 AFAC..... 4215
 1515 snAsnAspGIYAsnPhetValAsnAspLeuSerGIYArgArgGIYIleYr 1531
 4216 ...GCCGATTGGCTCAGATTTCGCAAAACCGCGAGTCCGGAATGGGG 4262

1532 GlnAlaGlyLeuYsaIaSerPheSerThrLeuSerGlyHisLeuGI 1548
 4263 CGTAAAGCCGCAATCAAGGTTTCAGCGGTGTCCTCCAGCGTCCGCC 4312
 1548 yval.....GlyTyrSer.....HisGlyAlaGly 1557
 4313 CCAAGGCCCG 4323
 1557 algiSerPro 1560

seq_name: p1r2:D85912

seq_documentation_block:
 hypothetical protein ypJ4 [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: D85912
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: D85912
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1528 <STO>
 A:Cross-references: GB:AE005174; NID:q12517083; PIND:ANG57760.1; GSPDB:GN00145; UWGP:Z35
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ypJ4

alignment_scores:
 Quality: 366.00 Length: 1689
 Ratio: 0.468 Gaps: 89
 Percent Similarity: 46.300 Percent Identity: 20.130

alignment_block:
 US-09-303-518D-649 x D85912 ..

Align seg 1/1 to: D85912 from: 1 to: 1528

268 ATGATGATTTTCTGTGTCGCTAACGCCGCGCGCATGCTGGG 317
 17 LeuIleSerAlaLeuIleTyrAlaProGlyMetAlaAlaPheThrPr 33
 318 CGATCAATATATTGTAGCGTGCACATAAC.....GGCGCTATA 358
 33 oasp.....ValIleGlyValAlaAsnSpoluhThrValaspGlySerG 48
 359 ACAACGTGATTTTGTGCGGGAAGCAAAATCCGATCAACATCGTTT 408
 48 lnaTyrValasp.....GluArgGlyThrThrAsnAsnThrHis..... 60
 409 ACTTAATAATTTGTGAACGAAATATATAAGCAGGAGCTAAAGCCCA 458
 61IleIleAsnHisGlyGlnGlnAsnVal..... 69
 459 TCCTATATGGCGCGATTCATATATCCGCTTTCATTAATTTGTACAG 508
 70TyrGlyGly.....ValSerA 75
 509 ATGCAGACCTGTTGAATGACCACTTATGAT...GGCGG..... 549
 75 snGlySerLeuIleGlnSerGlyTyrGlnAspValGlyArgHisAsn 91
 550 AAATATATGATCAAAATATATACCTGACGCTGTCATTTGGGCGAG 599
 92 AsnTyrValGlyGlnSerAsn.....AsnThrThrIleAsnGlyGI 105
 600 CAGGCAATATTTGGCATCTGATGAAGAT..... 627

105 yArgGln.....SerIleHisAspGlyGlyIleSerThrGlyThrI 119
 628GAGCCCAATTAACCGAAAGTTTCATATCATATTTGACGCTGAT 672
 119 leIleGlnSerGlyAsnGlnAspValTyrGlyGlyIleSerAsnGly 135
 673 TCTGGCTGCTGTGGCAATACCTTTGCACAAATATGATCAGGTGTGG 722
 136 ThrThrIleGlyGlyAlaSerArgValGlnGlyGlySerAlaAsnGI 152
 723 CACA...GTCACTTAGTAGTGAATAAATTAACATAGCCCATATGCTT 769
 152 yThrIleAspGlyGlySerGlnIleValAlaGlnGlyHisAla. 168
 770 TTTTACCAACAGGAGCGCTATTGGCGACAGTGCACCAATGTTTATC 819
 169AspGlyThrThrIleAsnGlySerGlySerGlnAspValAl 182
 820 TATGATGCCCAAAACCAAAAGCTTATTAATGCGGTATTCACAAACGG 869
 183 GlnGlySerLeuAlaThrAsnThrThrIleAsnGly.....GI 195
 870 CAACCCCTATATATAGCA.....AAAGCA 892
 195 yArgGlnTyrValGlnGlnSerThrValGlnThrThrIleGlyAsnG 212
 893 ATGCTTCACAGCTGCTGCTGAAGATTTCTATGATGAATCTTTGCT 942
 212 IyGlyGlnGlnIlaArgValTyrGlnSerArgAlaLeuAspThrThrIleGlu 228
 943 GGAGATACCCCATTCAGTATTCACGAACACGCTCAAAATGGGAAA..... 987
 229 GlyGlyThrGlnSerLeu.....AsnSerLysSerTh 239
 987 987
 239 lAlaLysAsnThrGlnIleTyrSerGlyGlyThrGlnIleIleAspAsnT 256
 988TACTTTTAAAC.....GACGAT 1005
 256 hrSerSerSerAspValIleGluValTyrSerGlyGlyValLeuAspVal 272
 1006 AATATGCGCAGAGAAAATCAATGCCAACAACACCAATCTCTGCC 1055
 273 SerGlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaIleLeu 289
 1056 TAATAGATTAAAAAACGAAACGCTTCAATGTTTAATGTTTCTTATCCG 1105
 289 sThrAsnThrAsnGlyThrThrValSerGlyThrAsn.....Serg 303
 1106 AGACAGCAGAGAACCTGTTATTCATGCTGCA..... 1137
 303 lGlnGlyAlaPheSerIleHisAsnHisValAlaAspAsnValLeuGlu 319
 1138 ...GGTGT.....GTCAACAGTTAT..... 1155
 320 AsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerAlaAsnLysThrII 336
 1156C 1156
 336 eIleLysAspLysGlyThrMetSerValLeuThrAsnAlaLysAlaSpA 353
 1157 GACCCAGACTGAATTAATGCA..... 1176
 353 lAThrArgIleAspAsnGlyGlyValMetAspValAlaGlyAsnAlaThr 369
 1177GAAATATTTCTTATTTATGACGAAGC 1202
 370 AsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnAsnTyrGlyLeaI 386
 1203 AAAAGCGAATGATGATCTTACGACGAACATCATCAAGT..... 1242
 386 aThrGly.....ThrAsnIleAsnSerGlyThrGlnAsnI 398

1243CCTGGAGATATATATTCAGAGAGATTTTACGCTCGCTCGAA 1287
1288 AATAAC.....GAACTTGGCAAGCGCGCGCTTCAT 1322
1289 leysserglyly.....lysalaasprthrthlleiser 411
412 glyserarglnvalvalglulysapglthrthalleglyserasn 428
1323 CAGTGAAGACAGTACCGCTTACTTGAAAGTAAACGGCGTGGCAACAGC 1371
428 eseralaglyglyserleuilevaltythrlyglylethahisgly 445
1372CGCTGTCCAAATTCGGCAAGCGCAGC 1398
445 alasnglnluthrlyseralaleuvalalaasnthrlyalaglythr 461
1399CTGCAGCTTCAGCGCAAGG..... 1419
462 asplieglylytyrasnlyleuserhithrthlethrlyglyl 478
1420GAAACCAAGCTCGATCAGCTGGCGCAGC 1450
478 ualasnthyralvalleugluasnthrlygluethrvalalal 495
1451 GTACAGTATTTGGATCAGCAGCGAGTAAAGCGAA..... 1491
495 ysthseralalysasnthrthlleasplaglyglyleuileval 511
1492AACAAGCCTTTAGTAAATCGCTGGTTCAGCGCGAGGTACGCT 1538
512 glnlysglnalalythrAspserthrargleusnasnlyglyvalle 528
1539 GCAACTGATCCGAT.....AATCAGTTCACACCGC 1570
528 uelvalaglnasplglyglualalalyshisvalgluinsnglsergly 545
1571 ACAAGATATTCGCTTCGCGGAGGAGCTTG.....GATTTAAC 1614
545 lyalaleuilealaserthrthrserythrleuileglulthrthsn 561
1615 GGGCATTCGCTTCGTCACCGATTCACAAATACCGATGAAGGGCGAT 1664
562 serytyglyAspalaphetyr...IlearganserGluAlalyasvsa 577
1665 GATTGTCAACCAATCAAGAAAGATCCACCTTACCTACAGGCA 1714
577 lvalleu.....Gluasnalaglyserleuthrvalalvalthrlyls 591
1715 ATAAAGATATTCGTACACCGGCAATACACAGCTTGATGCAAAAA 1764
591 etargvalavalasprthlleleasnalaasn.....Glyysmet 604
1765 GAATTCCTTACMACGCTTGTTGGCGGAGAAAT.....AGCAACCA 1808
605 Aspvat.....tyrily...Lysaspvalglythrvalle 615
1809 AACGAACGGCGGCTCAACCTGTTCACAGCGCGCGAGAAAGCGCA 1858
615 unsnseralaglythrthllethrlyalaseralathrseraplysa 632
1859 CCCTGTGCTTCGCGGAGAACAAATTAACGCAACATCACGCAACA 1908
632 laasnlelysglyglythrthvaltyrlyglyleuathrthglnala 648
1909 AAC.....GCAAACTGTTTTTCAGCGGAGACCAACACCGCAGC 1949
649 Asnllieglyserglygluinslvalasplglyserthrlyds.. 664
1950 CTACATTCATTTAAAGC..... 1968
665ThrHisleasnlyglythrthvalgluasnthyrglylsa 680

1969CATTCGCGCAAAAGAGGC 1989
680 lalleasnthrspillevalserglyleuglnlillemetalaasnly 696
1990 ATTCCTCGGGGAAATCGTG...TGGGACAAAGACTGGATCACCGCAC 2036
697 ThrAlaglyserlleleasnlyglyserGlnlevalasnglyl 713
2037 ATTTAAAGGGAACCTTCCAATTAAGCGGACAGCGGTGTTC 2086
713 yglyleuAlagluasnservalleusnaspolyglythrleuaspvala 730
2087 GCATGTTCGCAAGTGAAGCGATTCGATTCAGCAATCACGCCCA 2136
730 rglylulysglyseralathrgly.....lleclnlnsersergln 743
2137 GCAGTTTTGGTGTGCGCACCGCATCAACACCAATCTGCACGCTTC 2186
744 GlyAlaleuvalAlathrthraralathrargvalthrlythrarval 760
2187 GCACTGAGCGGCTGCAAAATGTGCGAAAA..... 2220
760 aasp.....glyvalAlapheserlleglulnglyAlalaasnani 775
2221ACCATTCAGCAGAT..... 2235
775 leleuenualaasnlyglyvalleuthrvalgluuserapthrsSer 791
2236AAGTGTGC 2246
792 AsplysThrGlnValasnthrlyglyargglullevallysthrlysa 808
2247 TTCATTGACTTAAGACGACATCAGCGCAATGTCATCT..... 2286
808 arthrAlathrglyThrthleuthrlyglygluinslvalleugly 825
2287 .GCCATCAGCTCATTTAAATCTCACAGGCTTCACACACACAGCGC 2334
825 alAlaasnlythrthlleasnaspolyglylleclnlnthr..... 838
2335 AATCTTATGCAATGCGCATACA...CGTTATACAGTACGCGCACAGC 2381
839 ...ValserAlaasnlyglualalelysthrthlleasnlyglyl 854
2382 CACC.....CAAAAGCGCAACTTAC...CTCGTGGCAATG 2416
854 ythrleuthrValasnspasnlylsalathrAspllevalGlnasns 871
2417 CCCAAGCAATTTAATCAAGCCACATTAAGCGCAACATCGGCTGC 2466
871 erglyAlalaaleuglnthrSerthralasnly...llegluileser 886
2467 GGCAT.....GCTTCATTTAATCTAAGCGACCGCGCTACAAA 2507
887 glythrhisglnlyrglythrpheserlleSerGlyasnleuathrAs 903
2508 CGGCACTGACGCTTCGCGCAACGCTAAGCAACGTAAGCATTCG 2557
903 nmetleuengluasnlyglyasnleuvalleuAlaglythrclua 920
2558 CACTCAACGGTAATCTCCTACCGCATTAAGCAGATTCATTTGAA 2607
920 laargasprthrvalglylysgly...AlawetGlnasnngly 935
2608 AGCAGCGCTTACCGAGCAAAATCAGCGGCGCAAGATACGCGCATTA 2657
936 Glnaspralathrlyvalasnserglylyglthrlythreuglyar 952
2658 CTAAAGAC.....AGCAATGCAGCTGC 2683
952 gserlyasplubheglnala**AlaargAlagluasnleuglnvala 969
2684 CGTCAGCGCAGGAATTA..... 2700

969 IAGLYGTYHrAlaLeValTYrAlaGlyThrLeuAlaAspAlaSerVal 985
2701GGCAATTAAACCTTGACACGCCAC 2726
986 SerGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAlaValTh 1002
2727 C...ATTACACTCAATTCCGCTATCGC...CACGATCGCGCAGGGCGC 2770
1002 rProValIysLeuGluGlyAlaLeArgIleThrAspSerAlaThrLeu 1019
2771 AAACGGGAGTGGCAGATGCGCGCGCGCTTGCGCGCTTCGCGC 2820
1019 hrIleGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSer 1035
2821 CGT 2823
1036 ArgGlySerValThrLeuAsnSerAsnAsnSerCysAlaGlyThrSerAs 1052
2824TCCCTATTA 2832
1052 nCysGluTYrArgValAsnSerLeuLeuAsnAspGlyAsnValTYrL 1069
2833TCCGTTACACGCCCAACTCGGTAGATCCGTTTCACACG 2874
1069 euserAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTYrAsnThr 1085
2875 CTGACGTAACGCGCAATTGAACGCTCAGGAACATTCGCTTTATGTC 2924
1086 LeuThrThrAsn...GluLeuSerGlySerGlyAsnPhenylLeuHisTh 1101
2925 GGAACCTTGGGCTACCGCAGCCACAAATGAACCTGGCGGAAAGTTCCG 2974
1101 rAsnValAlaGlySerArgIlyAspGlnLeuValValAsnAsnAlaAT 1118
2975 AAGGCACTTACACCTTGGCGGTACACATACGCGCAACGAACCTGGAAC 3024
1118 hrGlyAsnPhenylIlePheValGlnAspThrGlyValSerProGlnSer 1134
3025 CTGGAACAATTGACGCTAGTGAAGGAAAGACAAACCGCTTCGCA 3074
1135 AspAspAlaMetThrLeuValLysThrGlyGlyL 1146
3075 AAACCTTAATTTACCCCTGCAGAAAGAACACGTCGATGCCGCGC 3118
1146 yAspAlaSerPheSerLeuGlyAsnThrGlyPheValAspLeuGly 1163
3119 CGTGGCGTTACCAACTCATCCGCAAGACGCGGAGTTCCGCTG 3162
1163 hrTYrGluTYr...ValLeuLysSerAspGlyAsnSerAsnThrAsnLeu 1178
3163 CATAAATCCGTCAGAAAGAAAGAGCTTTCGACAAACTCGCGCAAGGCAGA 3212
1179 ThrAsnAspAllylPro...AsnProAspProAsnProAsnProAsnPro 1195
3213 AGCCAAAACAGGCGGAGAAAGACAAACGCGAAAGCCTTGACGGGCTGA 3262
1195 snProLys 1197
3263 TTGCGCGCGGCGCATGCTCGCAAAAAGACAGAAAGCCTTGCGAAGCG 3312
1198ProAsp... 1199
3313 GCCCGCAGCAGAGGGGAGAAATGTCGATTAATGACGCGGAGAGAAAGA 3362
1199 1199
3363 GAAAAACGCGTGACAGCGGATTAAGACACGCTTGCGGAAACAGCGCG 3412
1199 1199
3413 AAGCGAAGAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCGC 3462
1200ProLysProAsp...ProLysProAspProL 1209
3463 CGCCGGGATTTGCCCACTGCAACCCCAACCCGACCGCCACCGGAGCG 3512
1209 yspPro...Asp...ProThrProGluProThrProThrProValProGlu 1224
3513 CGACCTGATCAGCGCTTATGCCAATAGCGGTTTGATGTAATTTCCGCA 3562
1224 sArgIleThrPro...SerThrAlaValLeuAsnMetAlaAlaAT 1239
3563 CGCTCAACAGCGTTTCGCCGTACAGACCAATTA 3597
1239 hrLeuProLeuValAlaPheAspAlaGluLeuAsnSerIleArgGluArgLeu 1255
3598 GACCGCGTATTTGCCAAGACCGCGCAACGCGCGCTTGACAAAGCGCAT 3647
1256 AsnIleMetLysAlaSerProHisAsnAsnValTTP...GlyAlaTh 1271
3648 CCGGGACACCAAAACACTACCGTTCCGACAGATTTCGCGCTTACCGCAAC 3697
1271 rTYrAsnThrArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGlu 1288
3698 AAACCGACCTGCGCCAAATCGSTATGCAGAAAAACCTC 3735
1288 LThrLeuThrGlyMetThrValGlyIleAspSerProAsn 1301
3736GCCAGCGGCGCGTCCGATCCTGTTTGCACAAACG 3773
1302 AspIleProGluGlyIleAlaThrLeuGlyAlaPheMetGlyTYrSerH 1318
3774 GACGAAAAACACTTCGACGACGCG 3798
1318 serHisIleGlyPheAspArgGlyGlyHisGlySerValGlySerTYrS 1335
3799 ..ATCGCAACTCGCGCGCTTCCGACGCGCGCTTTC 3837
1335 erLeuGlyGlyTYrAlaSerTrpGluHisGluSerGlyPheTYrLeuAsp 1351
3838 GGGCAATACGGCATGACAGGTCTACATCGCATACGCGC 3878
1352 GlyValValLysLeuAsnArgPheGluSerAsnValAlaGlyLysMetSe 1368
3879 GGGCGCGGGTTTACGACGCGCAGCTT...TCAAGCGCAATCGGAGCA 3925
1368 rSerGlyGlyAlaAlaAsnGlySerTYrHisSerAsnGlyLeuGlyLys 1385
3926 AAATCGCGCGCGCTGCTGCATTACGGCATTCAGGACGATACCGCGCC 3975
1385 sIleGluThrGlyMetArgPhe 1392
3976 GGTTCGCGGATTCGCGATCGAACCGGACATCGCGCAACGCGCTATT 4025
1393 ThrAspGlyAsnTrpAsnLeuThrProTYrAlaSerLeuThrGlyPheTh 1409
4026 CGTCAAAAACGAGATTACCGCTACGAAAACGTCATA 4063
1409 rAlaAspAsnProGluTYrHisLeuSerAsnGlyMetGluSerLysServ 1426
4064 TCGCCACCGCGCGCTTGATTCACACCGCTTACCGC...GGGCGCATTAAG 4110
1426 AlaThrThrArgSerIleTYrArgGluLeuGlyAlaThr 1438
4111 GCAGATTATTCATTCAACCGCGCAACGCGCATTCGATCCATCGCTTATT 4160
1439 LeuSerTYrAsnMetArgLeuGlyAsnGlyMetGluIleGluProTyrPle 1455
4161 GAGCTGTCTATACGATGCGCTTGCGCAAAAGTCCGAACAGCGGTCA 4210
1455 uLysAlaAlaValArgLysGluPheValAspAspAsnArgValLysVal 1472
4211 ATAC 4215
1472 snAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArgGlyIleTYr 1488

```

4216 ...GCCGATATGGCTCAGGATTTGCGCAAAACCGGACGCGGAATGGG 4262
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
1489 GlnIaGlyIleLysAlaSerPheSerThrLeuSerGlyHisLeuGI 1505
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
4263 CGTAAAGCCGCAATCAAGTTTCACGCTGTCCCTCCAGCGTGGCCG 4312
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
1505 yVal.....GlyTyrSer.....HisGlyAlaGly 1514

4313 CCAAGGCCCG 4323
      ::::::::::|||
1514 algiuserpro 1517

seq_name: pir2:A65044

seq_documentation_block:
hypothetical protein b2647 - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C/Accession: A65044
R/Blattnier, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; M01D:97426617
A/Accession: A65044
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1569 <BLAST>
A/Cross-references: GB:AE000350; GB:U00096; NID:g2367147; PIDN:AAC75695.1; PID:g1789001;
A/Experimental source: strain K-12, substrain MG1655

alignment_scores:
Quality: 364.50 Length: 1634
Ratio: 0.485 Gaps: 81
Percent Similarity: 46.022 Percent Identity: 19.951

alignment_block:
US-09-303-518d-649 x A65044 ..

Align seg 1/1 to: A65044 from: 1 to: 1569

445 GGGAGTAAGGCCATCCTATGCGCGGATATCATGATCGCGCTTTCGA 494
      |||||  ::  ::|||::|||  |||::|||  |||::|||  |||
97 GlyThrThrAsnThrHisIleLeuAsnHisGlyGlnGlnAsnValHi 113
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
495 TAAATTTGTACAGATGCGAAGACCTGTGAAATGACAGTATATGAT 543
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
113 sGlyGlyValSerAsnGlySerLeuIleGlySerGlyGlyTyrGlnAspI 130
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
544 .....GGCGGCAATATATCATGCAAAATTAATTACCGTACCGGTGT 585
      ::::::::::|||  |||  |||  |||  |||  |||  |||
130 leGlySerHisAsnAsnPheValGlyGlnAlaAsn.....AsnThr 143
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
586 CGTATGGGGGAGCGAGCAATATGCGCATCTGATGAAGATGACCCAA 635
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
144 ThrIleAsnGlyIleArgGln..... 150
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
636 TAAACCGGAAGTTTCATATCATATTGCA.....AGTGGTATTTCTGGC 679
      |||  |||  |||  ::  ::|||::|||  |||::|||  |||
151 .....SerIleHisAspGlyIleSerThrGlyThrThrI 163
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
680 TCGTTGTGGCAATACCTTTCGACAAATGGA.....TCAGTGGTGGC 723
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
163 leGlySerGlyAsnGlnAspValTyrGlyGlyIleSerAsnGlyThr 179
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
724 ACAGTCACTAGTAGTGAATAAATTA.....CATAGCCCATATGCTTT 770
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
180 ThrIleGlyGlyIleAlaSerArgValGlyGlySerAlaAsnGlyI 196
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
771 TTTCACACAGAGGCTCA..... 789
      |||  |||  |||  |||  |||  |||  |||

```

```

196 eleuIleAspGlyGlySerGlnIleValLysValGlnGlyHisAlaAsp 213
790 .....TTTGGCAGACGTGGCTCCACCAATGTTATGATGAGCCCAA 831
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
213 LyThrThrIleAsnLysSerGlySerGlnAspValValGlnGlySerLeu 229
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
832 AAGCAAAAGTGGTAAATTAATGGGGTATTCGCAAGGCGCAACCCCTATAT 881
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
230 AlaThrAsnThrThrIleAsnGly.....GlyArgGlnTyrVal 242
882 AGCA.....AAAGCATGGCTCCAGC 904
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
242 IeGlnSerThrValGluThrThrThrIleLysAsnGlyGlyGlnAla 259
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
905 TGGTTCGTAAGATGGTCTCTATGTAATCTTGTCTGAGATACCAT 954
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
259 rGlyIleGlySerArgAlaLeuAspThrThrIleGlyGlyGlyThrGln 275
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
955 TCAGTATTTACGACGACGCAAAATGGGAAATCTTTTAAACGAC.. 1002
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
276 SerLeuAsnSerLysSerThrAlaLysAsnThrHisIleTyrSerGly 292
1003 .....GATTAATATGGCGACAGCAAAATTCATGGCAAAACATG 1039
292 yThrGlnIleValAlaAspAsnThrSerThrSerAspValIleGluVal 309
1040 AACCAATCTCTGCTAATACA..... 1062
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
309 eGlyGlyValLeuAspValArgGlyGlyThrAlaThrAsnValThrGln 325
1063 .....TTAAACACGACCCGTTCAATGTTTAAAGTTTC 1097
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
326 HisAspGlyAlaIleLeuLysThrAsnThr.....AsnGlyThrThr 339
1098 TTTATCCGAGACGACGCAAGAACCTGTTAT.....CATGCTG 1135
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
339 rValSerGlyThrAsnSerGlyValAlaPheSerIleHisAsnHisVal 356
1136 CA.....GGTGGT.....GTCAACAGTTAT 1155
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
356 IAspAsnValLeuLeuGluAsnGlyGlyHisLeuAspIleAsnAlaTyr 372
1155 ..... 1155
373 GlySerAlaAsnLysThrIleIleLysAspLysGlyThrMetSerVal 389
1156 .....CGACCCAGACTGAATATGCA..... 1176
389 uThrAsnAlaLysAlaAspAlaThrArgIleAspAsnGlyGlyValMet 406
1177 .....GAAAT 1182
406 sPValAlaGlyAsnAlaThrAsnThrIleIleAsnGlyGlyThrGlnAsn 422
1183 ATTTCCTTTATTCGCAAGGAAAGCGCAATGTACTTACCAACAAT 1232
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
423 IleAsnAsnTyrGlyIleAlaThrGly.....ThrAsnI 424
1233 CAATCAAGT.....GCTGAGGATTAATTCCACAGAG 1267
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
434 eAsnSerGlyThrGlnAsnIleLysSerGlyGly.....LysAla 448
1268 ATTTTACGGCTCCCTGCAAAATTAAC.....GAAACTTG 1302
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
448 sPThrThrIleIleSerSerLysThrArgIleValValGlyLysAspGly 464
1303 CAAGCGCGGCGCTTCATATCATGAGTGAAGACATCCCTTACTTGGAAGT 1352
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
465 ThrAlaIleGlySerAsnIleSerAlaGlyGlySerLeuIleValTyr 481
1353 AACGGCGTGGCAACGAC.....CGCTGT 1378
      |||  ::  ::|||::|||  |||::|||  |||::|||  |||
481 rGlyGlyIleAlaHisGlyValaAsnGlnGluThrGlySerAlaLeuVal 498

```

```

1379 CCAAAATCGCAAGGACG.....CTGCAC 1404
      ::::: ||| |||||
498 laasnthrglyalaglythrAspIleuglyTyrAsnlyLeuSerHis 514
1405 GTTCAAGCCAAAGG.....GAAACCAAG 1430
      |||
515 PheThrIleThrIleGlyGlyGluAlaAsnTyrValValLeuGluAsnThrG1 531
      ::::: ||| |||
1431 CTCGATCAGGTGGCGGACGCTACGTATTTGGATCAGCAGGACGAC 1480
      ::::: ||| |||
531 ygluLeuThrValValAlaLalysThrSerAlaLysAsnThrThrIleAsp 548
      ::::: ||| |||
1481 ATAAAGGCAAA.....AAACAAGCCTTTAGTCAAAATCGGCTTG 1518
      |||||
548 hrGlyGlyLysLeuIleValGluLysGluAlaLysThrAspSerThrArg 564
1519 GTCAAGGCGGCGGTACGCTGCTGCACTGAATGCCGAT..... 1554
      ::::: ||| |||||
565 LeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAlaLysH1 581
1555 ....AATCAGTTCAACCCGACAACTCTATTTCGGCTTCCGGCGGAC 1600
      ::::: ||| |||
581 sValGluGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSerGlyT 598
1601 GTTTC.....GATTTAAAGGGCATTCGCTTCCGTCACCGTATTCA 1644
      |||
598 hrLeuIleGluGlyThrAsnSerTyrGlyAspAlaPheTyr...IleArg 613
1645 AATACGATGAAGGGCGATGATGTCACCAACCAATCAAGACAAAGAATC 1694
      |||||
614 AsnSerGluAlaLysAsnValValLeu.....GluAsnAlaGlySe 627
1695 CACCGTTACCATTTACAGGCATTAAGATATGCTACACCGGCAATACA 1744
      |||
627 rLeuThrValValThrGlySerArgAlaValAspThrIleIleAsnAla 644
1745 ACAGCTTGATAGCAAAAAAGAAATGGCTACACGCTTGCTGGCGGAG 1794
      |||
644 sn.....GlyPysMetAspVal.....TyrGly... 651
1795 AAAGAT.....ACGACCAAAAGAGCGGCGCTCAACCTTGTTACCA 1838
      |||||
652 LysAspValGlyThrValLeuAsnSerIleGlyThrGlnThrIleTyrAl 668
1839 GCCCGCGGAGAAAGACCGCAGCTCTCTCTCCGCGGCAAAATTTAA 1888
      |||||
668 aserIleThrSerAspLysAlaAsnIleLysGlyGlyLysGlnThrValT 685
1889 ACGGCACATCCAGCAACAAC.....GGCAAACTGTTTTCAGC 1929
      |||
685 yrcGlyLeuAlaThrGluAlaAsnIleGlnSerGlyGluGlnIleValAsp 701
1930 GGCAGACCAACACGCGCAGCTTACATATTAAACGAC..... 1968
      |||
702 GlyGlySerThrGlnLys.....ThrHisLeuAsnGlyGlyThrGlnTh 716
1969 .....C 1969
716 rValGlnAsnTyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuG 733
1970 ATTGGTCGCAAAAGAGGCGATTCCTCGCGGGAATCGTG...TGGGAC 2016
      ::::: ||| |||||
733 lngInIleMetAlaAsnGlyThrAlaGluGlySerIleIleAsnGlyGly 749
2017 AACGATCGATCAACCGCACAATTAAAGCGAAACTTCCAAATTTAAAG 2066
      ::::: ||| |||||
750 SerGlnValValAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnS 766
2067 CGGACAGCGCGGTGGTTCCGCAATTTGGCAAAAGTAAAGCGGATGGC 2116
      |||
766 pGlyGlyThrLeuAspValArgGlyLysGlySerAlaThrGly..... 780

```

```

2117 ATTGAGCAATCAAGCCCAAGCAGTTTGTGTGTCAGCCGATCAAGC 2166
      ::::: ||| |||||
781 ..IleGlnGlnSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThr 796
2167 CACGACATCTGTACAGCTTCGCGACTGCGAGGCTGCACAAATTTGTGCA 2216
      |||||
797 ArgValThrGlyThrArgAlaAsp.....GlyValAlaPheSerIleG1 811
2217 AAAA.....ACCA 2224
      |||
811 uGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyValLeuThrIV 828
2225 TTACCGACAT..... 2235
      ::::: ||| |||
828 alGlnSerAspThrSerSerAspLysThrGlnValAsnMetGlyGlyArg 844
2236 .....AAAGTATGCTTCATTTAGTAAAGCCGACATCAGCGCA 2276
      |||||
845 GluIleValLysThrLysAlaThrAlaThrGlyThrThrLeuThrGlyG1 861
2277 TGTGATCTT.....GCCGATCCGCTCATTTAAATCTCACAG 2314
      |||
861 yGluGlnIleValGluGlyValAlaAsnGlnThrThrIleAsnAspGlyG 878
2315 GGCTTGCCACACTCAACGGCATCTTAGTGCAAATGGCGAT..... 2355
      |||||
878 LysIleGlnThr.....ValSerAlaAsnGlyGluAlaIleLys 890
2356 ACAGCTTATACAGTACGACCAACGCCACCCCAAGCGCAAC..... 2397
      |||||
891 ThrLysIleAsnGluGlyGlyThrLeuThrValAsnAspAsnGlyLysAl 907
2398 ..CTTACCTTCGTGGCAATGCCCAAGACAACTTAATCAACGCAATTA 2446
      |||||
907 arThrIleValGluIleAsnSerGlyAlaAlaLeuGlnThrSerThrAla 924
2447 ACGGCACACATCGGCTTCGGGCAT.....GCTTCATTTAACTA 2487
      |||||
924 smGly...IleGluIleSerGlyThrHisGlnTyrGlyThrPheSerIle 939
2488 ACGGACACGCGCGTCAAAAGCGGAGTCTGACGCTTCCGCAACGCTAA 2537
      |||
940 SerGlyAsnLeuAlaThrAsnMetLeuLeuGlnGlnGlyAsnLeuLe 956
2538 GGCAAACGTAAGCATTCGCGCATCAAGCTCAAGTAAATCTCCCTACGCATA 2587
      |||||
956 uValLeuAlaGlyThrGluAlaArgAspSerThrValGlyLysGlyGly 972
2588 AGCGATATTCATTTTGAAGCAGCGCTTACCGGCAAAATCAGCGGC 2637
      |||||
973 ..AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsnSerGly 988
2638 GGCAGATATACGGCATTTACACTTAAAGACGCGAATGACGCTGCCGTC 2687
      |||||
988 GlyGlnTyrThrLeuGlyArgSerLysAspLupheGlnAlaLeuAlaAr 1005
2688 AGCGACGAATTA..... 2700
      ::::: ||| |||
1005 gAlaGlnAspLeuGlnValAlaGlyGlyThrAlaIleValTyrAlaGlyT 1022
2701 .....GGCAATTTAACTT... 2715
      |||||
1022 hrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSerLeuMet 1038
2716 .....GACACGGCACCC..ATTACACTCAATTCGCGCTATGCG... 2751
      |||||
1039 ThrProArgAspAsnValThrProValGluGlnGlyAlaValArgI1 1055
2752 ..CACGATCGCGGCGGCGCAACCGGCAAGTGCAGAGATGCGCGCGCG 2800
      |||||
1055 eThrAspSerAlaThrLeuThrLeuGlyAsnGlyValAspThrThrLeuA 1072
2801 GCCGTTGCGCGCGTTCGCGCGGT..... 2823

```



```

1072 laaspleuthrAlaalaSerArglySerValTrpleuAsnSerAsn 1088
      :::: |||
2824 ..... TCCCTATTA.. 2832
      |||||
1089 SerCysAlaGlyThrSerAsnCysGlyTrArgValAsnSerleu 1105
      |||||
2833 ..... TCCGTTACCGCGCACTTCGG 2854
      ::::: |||||
1105 uAsnAspGlyAspValTrleuSerAlaGlnThrAlaAlaProAlaThrT 1122
      ::::: |||||
2855 TACAATCCCTTCAACAGCGGTAAAGCGCAATGACGGTCAG 2904
      ::::: |||||
1122 hrasnglyIleTrAsnThrleuThrAsn...GluLeuSerGlySer 1137
      ::::: |||||
2905 GAACATTCCTTATGTGGAACCTTCGCTACCGCAGCAAAAT 2954
      ::::: |||||
1138 GlyAsnThrleuHisThrAsnValAlaGlySerArglyAspGln 1154
      ::::: |||||
2955 GAACTGCGGAAAGTTCCGAAAGGCACTTACACCTGGCGTCAACAATA 3004
      ::::: |||||
1154 uValValAsnAsnAlaThrGlyAsnPheLysIlePheValGlnspt 1171
      ::::: |||||
3005 CCGGCAACGAACCTCGAAGCCTCGAACATTTGAGTGAAGGAAA 3054
      |||||
1171 hrGlyValSerProGlnSerAspAspAlaMetThrleuVal..... 1184
      ::::: |||||
3055 GACAACAACCGCTGTCCGAAACCTTAATTTACCTCGCAAAAGCA.. 3102
      ::::: |||||
1185 ..... LysThrGlyGlyGlyAspAlaSerPheThrleuGlyAsnThrG 1199
      ::::: |||||
3103 .... CAGGTGATGCGGCGGCGGTTCACATCATCGCAAAAGC 3148
      |||||
1199 yGlyPheAlaAspLeuGlyThrTrpGlyTrp...ValLeuLysSerSp 1215
      ::::: |||||
3149 GCGAGTTCGCGCTCATTAATCCGCTCAAGAAAGACTTTCCGACAAA 3198
      |||||
1215 Ly.AsnSer..... 1217
      ::::: |||||
3199 CTGCGCAAGGAGAACCAAAAGCGGAAAGAACAAAGCGGCAAG 3248
      ::::: |||||
1218 ..... AsnTrpA 1220
      |||||
3249 CCTGACGCGCTGATTCGGCGCGCGGATCCGTCGAAAAGACAGAAA 3298
      |||||
1220 snleuThrAsnAspValLysProAsnProAspPro..... 1231
      ::::: |||||
3299 GCGTTGCCGAACCGGCGCGGCAAGCGGCGGAAATGTGCGCATTAATG 3348
      ::::: |||||
1232 ...IleProAsn..... 1234
      ::::: |||||
3349 CAGCGGAGGAGAAAGAAAGGCGTGCAGCGGATAAAGACCGGCTT 3398
      ::::: |||||
1234 ..... 1234
      |||||
3399 GCGCAACAGCGCGAAGCGGAAACCGCGCGATACACCGCTTCCCC 3448
      |||||
1235 ..... ProLysProAspProLysProAspProLysPro 1245
      ::::: |||||
3449 GCGCGCGCGCGCGCGCGGATTCGGCAACTGCACACCCCAACCGAG 3498
      ::::: |||||
1246 AspPro...AsnProLysProAsp...ProThrProAspProThrProThr 1260
      ::::: |||||
3499 CCGCAACGCGCAGCGCGCTGATCAGCGTTATGCAATAGCGGTTTGA 3548
      |||||
1261 ProValProGlnLysArgIleThrPro.....SerThrAlaAlaValle 1275
      ::::: |||||
3549 TGAATTTTCCGCGCTCAACAGCTTTC.....GCGC 3583
      ::::: |||||
1275 uAsnMetAlaAlaThrLeuProLeuValPheAspAlaGlnleuAsnSer 1292
      ::::: |||||
3584 TACAGAGCAATTAGACGCGGTATTTGCCGAGAACCGCGCAACCGCTT 3633
      ::::: |||||

```

```

1292 leArgGluArgLeuAsnIleMetLysAlaSerProHisAsnAsnVal 1308
      |||
3634 TGGACAGCGGCATCCGGGACACCAACACTACCGTTCCGAAATTTCCG 3683
      ::::: |||||
1309 Trp...GlyAlaThrTrpAsnThrArgAsnAsnValThrThrAspAlaG 1324
      ::::: |||||
3684 GCGCTACCGCCCAACAAACCGACCTGCGCCCAATGCGTATCGCAAAA 3733
      ::::: |||||
1324 yAlaGlyPheGlnGlnThr.....LeuThrGlyMetThrValGlyI 1338
      ::::: |||||
3734 TC.....GCGACGCGCGCGCTGCGCATCTG 3759
      ::::: |||||
1338 leAspSerArgAsnAspIleProGlnGlyIleThrThrLeuGlyAlaPhe 1354
      ::::: |||||
3760 TTTTGCACAAACCGGCAAAACACTTCGACGACGAGC..... 3798
      ::::: |||||
1355 MetGlyTrpSerHisSerHisIleGlyPheAspArgIleGlyHisGlyse 1371
      ::::: |||||
3799 .....ATCGGCACCTCGCACAGCGCTTCCGACGCGCGC 3832
      ::::: |||||
1371 rValGlySerTrpSerleuGlyGlyTrpAlaSerTrpGlnHisGlnSerG 1388
      ::::: |||||
3833 TTTTC.....GGCAGATACGCGCATCGACAGGTTTC.....TAC 3864
      ::::: |||||
1388 TrpThrTrpLeuAspGlyValValLysLeuAsnArgPheLysSerAsnVal 1404
      ::::: |||||
3865 ATCGCATACGCGCGCGCGGCTTTAGACGCGGACGCTT...TCAGA 3911
      ::::: |||||
1405 AlaGlyLysMetSerSerClyGlyAlaAlaAsnGlySerTrpHisSerAs 1421
      ::::: |||||
3912 CCGCATCGAGGCAAAATCCGCGCGCGCTGCTGATTAACGCAATTCAG 3961
      ::::: |||||
1421 nGlyLeuGlyGlyHis.....IleGlnT 1429
      ::::: |||||
3962 CAGCATACGCGCGCGGTTTCGCGGATTCGCGCATCGCAACGACATCCG 4011
      ::::: |||||
1429 hrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTrpLysLaser 1445
      ::::: |||||
4012 GCAACGCGCTATTTGTCGCAAAAGCGATTAACGCGCTACGAAAC.... 4056
      ::::: |||||
1446 leuThrGlyPheThrAlaAspAsnProGlnTrpHisLeuSerAsnGlyMe 1462
      ::::: |||||
4057 .....GTCAATATCGCCACCGCGCGCTTCATTCACACCGCTACCGC. 4098
      ::::: |||||
4099 ...GCGGCAATTAAAGCAGATTAATTCATTAACCGCGGCAACATTTCC 4146
      ::::: |||||
1462 LysSerLysSerValAspThrArgSerTrp.....TrpArgG 1475
      ::::: |||||
1475 IuLeuGlyAlaThrleuSerTrpAsnMetArgLeuGlyAsnGlyMetGln 1491
      ::::: |||||
4147 ATCAGCGCTTATTTAGCCTGTTCCTATACGATGCGCGTTGCGGCAAGT 4196
      ::::: |||||
1492 ValGlnProTrpLeuLysAlaAlaValArgLysGlnPheValAspAsp 1508
      ::::: |||||
4197 CCGACACGCGCATATACGCGCTATTTGGCTCAGATTTTC.....GGCA 4240
      ::::: |||||
1508 nArgValLysValAsnSerAspLysAsnPheValAsnTrpLysSerGlyA 1525
      ::::: |||||
4241 AAACCGCAGTGGCAATGGGCGGTAAAGCGCAAAATCAAGGTTTCAG 4290
      ::::: |||||
1525 rArgGlyIleTrpGlnAlaGlyIleLysAlaSerPheSerSer...Thr 1540
      ::::: |||||
4291 CTGTCCTCGAGCGTGGCGCGGCAAGGCGCCACACTGGAACGCAACA 4340
      ::::: |||||
1541 leuSerGlyHisLeuGlyValGly.....TySerH 1551
      ::::: |||||
4341 CAGCGCGCGCATCAAA.....TTAGCTACCGCTG 4371
      ::::: |||||
1551 sSerAlaGlyValGlnSerProTrpAsnAlaValAlaGlyValAsnTrp 1567
      ::::: |||||

```

seq_name: p1r2:B90674
seq_documentation_block:

Aldh-1 adhesion-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90674
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Ysunaga, T.; Kihara, S.; Shiba, T.; Hattoni, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: B90674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1327 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033785.1; PID:G13359819; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS0362

alignment_scores:

Quality:	325.00	Length:	1474
Ratio:	0.509	Gaps:	72
Percent Similarity:	43.284	Percent identity:	20.692

alignment_block:

US-09-303-518D-649 x B90674 ..

Align seg 1/1 to: B90674 from: 1 to: 1327

```
712 TCAGTGTGTCGACAGTCACCTAGTAAATTAACATAGCCC 761
|||||
96 SerGlyGlyGlyLeuAspIleThrGlyGlyValThrIleAsnGlyPr 112
762 ATATGTTTATACCAACAGAGGCTCATTT.....GGCGACATG 802
|||||
112 OleuAsnPhleuLeuLysGlyThrGlyPheLeuAsnValSerSnaIag 129
803 GGTCCACCATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAAT 852
|||||
129 LysGlyLeuThrAlaAspAspLeuThrGlySer.....Asn 141
853 GGGGTATTGCAACGGCAACCCCTATATAGAAAAGCATGGCTTCCA 902
|||||
142 SerGlyMetArgHisAspArgGlyThrPheAsnValSerAsnGlyGly 158
903 GCTG..... 906
158 sIleHisValLysGlyThrSerArgLeuThrTyLeuGlnGlyAsnValS 175
907 ..... GTTGTAAAGTTGGTCTATGATGAATCTTT 939
175 ergLysGlnGlySerGlnValAsnSerGlnThrPheMetGlyValTyr 191
940 GCTGAGATACCATTCAGTATCTACGAACCATGCAAAATGGGAATA 989
192 .....GlySerTy 194
990 CTCCTTTTAC.....GACGTAATATGCGACAGAAATCATG 1030
|||||
194 rGlyGlyAsnGlnTyrLeuSerValAsnAsnGly...GlyGlnValAsn 210
1031 CCAAACTAGAACACAAAT..... 1047
210 laArgLysGlnIleSerLeuGlyTyrTyrAspGlnValSerAspThrThr 226
1048 .....TCTGTCCCTAATAGATTAAACAGAACCTGTAATG.. 1086
227 LeuAlaValSerGlnGlyGlyLysLeSerAlaProThrIleSerLeuSe 243
1087 ..TTAATGTTCTTATCCGAGACAGAACAGAACTGTTATCAGCTG 1135
|||||
243 rThrAsnSerGlnLeuAlaLeuGlnGlySerAlaAlaLysA 260
```

```
1136 CAGTGTGTCGACAGTCTATTCAGCCAGACTGAATAAGCAAAATATT 1185
|||||
260 laAlaGlyIle.....IleAspIleGlyLysIle 269
1186 TCGTTTATTCAGCAAGAAAGCGAATGATCTATTCACAGAAC..... 1230
270 GluPheValThrAlaLysThrSerGlyLysIleThrLeuAsnHisTh 286
1230 ..... 1230
286 rAspLysAspAlaThrIleSerAlaAspIleValSerGlySerGlyL 303
1231 .....ATCAATCAAGTCTGGAGATTTATTTCCAGAGATTTT 1272
303 euGlyTyrIleAsnAlaLeuAsnGlyThrThrTyLeuThrGlyAsp... 318
1273 ACGTCTCGCTGAAATACGAACCTTGGCAAGCGCGGCTTCATAT 1322
319 .....AsnSerAlaPheSerGlyLys...ValLysI 328
1323 CAGTGAAGACAGTACCGTT.....ACTTGAAG 1351
328 eGlnGlnAsnGlyAlaLeuGlyIleThrGlnAsnIleGlyThrAlaG 345
1352 TAAACGGCGTGGCAAAACGACCCCTGTCCAAATCGCAAAAGCAGCTG 1401
345 IeAsn.....AsnArgGlyLysLeu 351
1402 CAGCTTCAAGCC.....AAAGGGAACCA 1427
352 HisLeuLysAlaAspAspSerMetThrPheAlaAsnLysIleSerGly 368
1428 AGGCTGATCAGCTGGCGGACGTCAGTCTTTGGATGACAGCGCAG 1477
368 nGlyThrIleSerIleAspSerGlyThrValGlnLeu..... 380
1478 ACGATAAAGGCAAAACAGCCTTACT...GAATCGCTGGTGCAGC 1524
381 .....ThrGlyAsnAsnThrAlaPheSerGlyTyrIleAspValAlaSer 395
1525 GGCAGG.....GTAAGGTGCAGT 1544
396 GlyAlaValAlaValIleSerGluAspLysAsnIleGlyAlaGlnLe 412
1545 GAATGCCGATATCAGTTC.....ACCCGACAACTCATTT 1582
412 uAspValAspGlyLysLeuGlnIleAsnAlaAsnLysAspThrValPhe 429
1583 TCGGCTTTCGGCGGCGAGCT.....TTGATTTTAAACGGGCAAT 1620
429 sPAsnAspLeuGlnGlyArgGlyIleValGlnIleAsnMetGlyAsnHis 445
1621 TCGCTTGTTCACCGCTATTCAAATACGAT.....GAAGGGCGGAT 1664
446 GluPheSerPheAspGluPheAlaTyrThrAspThrPheGlnGlySer 462
1665 GATTGTCAACCAATCAAGACAAAGAAATCCACCGTTACCATTCACAGCA 1714
462 uAlaPheGlnAsn..... 466
1715 ATAAAGATATTGCTACACCGGCAATACACAGCTTGATACCAAAAA 1764
467 .....ThrThrPheAsn.....LeuGlnLysAsnAla 475
1765 GAATTTGCTACACGCTTGTGTTGGCGAAGAAATACCAACAAAGAA 1814
476 GluPheLeuGlnLysGly.....GlyIleThrAlaGlyL 487
1815 CGGCGGCTCAACCTGTTTACCAAGCCCGCGAGAAAGACCGACCCGAC 1864
487 nGlySerLeuValThrValGlyLysGlyAlaHisSerIleSerThrLeu 504
1865 TGCCTTCCGGCGGCAACAAATTTAAACGGCAACATCAGC..... 1902
```

```

504 1yPheSerGlyGlyThrValAspPheGlyAlaLeuThrAlaGlyAlaGln 520
1903 CAACAAACGCGCAACGCTGTTTTCAGCGGACGACCAACACCGCAGCCCTA 1952
521 MetThrGlyGlyThrValAsnValSer..... 529
1953 CAATCATTTAAACGACCATTTGTCGCAAAAGAGGCGATTCTCCGCGGG 2002
530 .....LysThrLeuAspLeuArgGlyG 537
2003 AA.....ATCGTGGGACACAGCTGGATC..... 2028
537 LngLysValIleGlnValSerAspSerAspValValArgSerValSerArg 553
2029 .....AACCGCAC 2036
554 AspLeuAspSerAlaLeuSerLeuThrGluValAspAspGlyAsnSerThr 570
2037 ATTTAA.....GCGGAA 2050
570 rIleLysLeuValAspAlaGlnGlyAlaGluValLeuGlyAspAlaGly 587
2051 ACTTCCAAAAT.....AAAGCGGACAGCGGCTGTTCCCGCAATGTT 2094
587 snLeuGlnLeuGlnAspLysAsnGlyGlnIleLeuSerSerSerAlaGln 603
2095 GCCAAAGTAAAGCGATTGGCATTTAGACATGACGCCCAAGCATTTT 2144
604 ArgAspIleGln.....GlnAsnGlyGlnLysAlaAlaVal 615
2145 TGGTGGCCACCGCATCAACGACACACATGTTACAGCTTCGAGCTGA 2194
615 lGlyThrLysAspTyrArgLeu...ThrSerGlyValAsnAsnAspGlyL 631
2195 CG.....GGTCTGACAAATGT...GTGAAAAAACCATTAAC 2229
631 euTyrIleGlyTyrGlyLeuThrGlnLeuAspLeuHisAlaThrAspSer 647
2220 GACGATAAGGATGCTTCATGACTAGACGAC..... 2265
648 AspAlaLeuValLeuSerSerAsnGlyLysSerGlnAsnAlaAlaAspLe 664
2266 .....ATCAGCGGCAATGTCATCTTGCC..... 2289
664 uSerAlaLysIleThrGlnLysGlyAspLeuAlaPheSerSerGlnLysG 681
2290 .....GATCAGCGCTCATTTAAATCTGCA 2313
681 lGlnThrValSerLeuSerAsnLysAspAsnAspTyrThrGlyValThr 697
2314 .....GGGCTGGCCACTCAAC.....GGCA 2336
698 AspLeuArgSerGlyThrLeuLeuAsnAsnAspAsnValLeuGlyAs 714
2337 T.....CTAGTCAAATGGCG 2353
714 nThrHisGluLeuArgLeuAlaAlaGluThrGluLeuAspMetAsnGlyH 731
2354 ATACACGTTATACAGTCAAGCCCAACGCCCAACCAACGCAACTTACG 2403
731 lSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeuLeuSer 747
2404 CTCGTGGGCAATGCCCAACGACATTTATCAAGCCACATTTAAAGCGCA 2453
748 LeuAsnGlyGlySerLeuThrValThrAsnGlyGlyThrSerThrGlySe 764
2454 CACATCGGCTTCGGGCAATGTTCAATTAACTAAAGCAGACGCCGTAC 2503
764 rLeuThrGlySerGlyGluLeuAsn.....IleG 774
2504 AAAACGCGAGTCTGACGCTTCCGGCAACGCTAAGCAACGTAAGCCAT 2553

```

```

774 lGlyGlyThrLeuAspIleAlaGlyAsp.....Asn 784
2554 TCCGACATCAACGGTAATGCTCCCTAGCCGATAGCA.....GT 2594
785 SerAsnLeuThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuVal 801
2595 ATTCCATTTTGAAGACAGCCGCTTACCGGACAAATACGCGGCGCAAG 2644
801 lSerHisAlaGlnGly..... 806
2645 ATACGCGATTACCTTAAGACAGCGAATGACGCTCCCTCAGCGACG 2694
806 ..... 806
2695 GAATAGCGAATTTAACTTAC...AAGCGACCATTAACATCAATTC 2741
807 ...LeuGlySerAlaAsnValGlnAsnAsnGlyThrLeuAlaLeuAsnAs 822
2742 CGCCTATCGCCAGATGCGGCGAGG.....GCGCAACCGGCA 2779
822 nSerAlaGlnLysArgAlaAlaAlaSerValAsnTyrAlaLeuGlyGlyA 839
2780 GTGCGACAGATCGCGCGCGCGCTTCGCGCGCTTCGCGCGTCCCTA 2829
839 snLeuThrAsnAsnGly..... 844
2830 TTATCCGTTACACCGCACTTCGTTAGATCCGTTCAACACGCTGAC 2879
845 ...ThrLeuMetThrGlyMetSerGlyGlnGlnAlaGlyAsnValLeuVal 860
2880 GGTAAACGCAATTTGACGGT...CAGGACATTTCCGCTTATGTCGG 2926
860 lValLysGlyAsnTyrHisGlyAsnAsnGlyGlnLeuValMetAsnThrY 877
2927 AACTCTTCGCTACCCGACG.....GACAAATTGAAGCTGCGGAAAT 2970
877 alLeuAsnGlyAspAspSerValThrAspLysLeuValValGluGlyAsp 893
2971 TCCGAAGGCACTTACACCTTGGCGGTCAACATACCGGCAAGCAACCTGC 3020
894 ThrSerGlyThrThrAlaValThrValAsnAsnAlaGlyGlyThrGlyAl 910
3021 AACCTTC.....GAACAAATTGACGGTAGTGAAGAAAGCAACA 3061
910 alTyrThrLeuAsnGlyTleGluLeuIleHisValAspGlyLysSerGluG 927
3062 AACCGCTGTCGAAACCTTAATTCACCTGCAACAAACGACAGCTGCAT 3111
927 lGluPhe.....ValGlnAlaGlyArgIleVal 936
3112 GCGGCGCGCGGCGGCTTACCACTCATCCGCAAGAGGCGAGTCCGCT 3161
937 AlaglyAlaLysTyrAspTyrThrLeuAlaArgGlyGlnGly..... 949
3162 GCATTAATCCGTCAAAGAACAGAGCTTCCGACAAACTCCGACAGGACG 3211
949 ..... 949
3212 AAGCAAAAAACAGCGCGAAGAAAGACACGCGCAAGCCTTGACGGCTG 3261
950 .....AlaAsnSerGlyAsnTyrPyr 956
3262 ATTGCGGCGCGCGCGGATGCGCGTGAAGAAAGCAAGAAAGCTTCGCAAC 3311
957 LeuThrSerGlySerAspSerProGlu..... 965
3312 GCGCGGCGAGGAGCGCGGAAATGTGCGCATTTATGACAGCGGAGAGAG 3361
965 ..... 965
3362 AGAAAAACGGGTGCAGCGGATTAAGACACCGCTTGGGAAACAGCGC 3411
966 .....LeuGlnPro 968

```

```

3412 GAAGCGGAAACCGCGGCTACACCGCTTCCCGCGCGCGCGCGC 3461
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
969 GUprometPromePromePromePromePromePromeProme 981
3462 CCGCGCGGATTTCCGCACTGCAACCGCACCGCGCGCGCGCGC 3511
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
982 .....ProgluprometPromePromePromePromeProme 994
3512 GCGACCTG.....ATCAGCGCTTATGCCAATAGCGGTTTG 3546
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
994 roAspleuAsnValAspAsnAspLeuArgProgluAlaGlySerTyrIle 1010
3547 AGTGAATTTTCCGCACTGCAACCGCTTCCGCG.....GTACAGA 3590
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1011 AlaAsnLeuAlaAlaAla...AsnThrMetPheThrThrArgLeuHisGly 1026
3591 CGAATTA.....GACCGCGCTATTTCGCGAAGACCGCC 3622
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1026 uATGLeuGlyAsnThrTyrTyrThrAspMetValThrGlyGluGlnLysG 1043
3623 GCAACGCGCTTTGACACAGCGGATCCGGGACACCAACACACGCTTCG 3672
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1043 LcThrThrMetThrMetArg.....HisGlyGlyHisAsnLysTrp 1057
3673 CAAGATTTCCGCGCGCTACCGCAACACCGACCTGCGCAATCGGTAT 3722
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1058 ArgAsp..... 1059
3723 GCGAATAAACCTCGGACGCGCGCGCTGCGGATCTGTTTCCGCAACAC 3772
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1060 .....GlySerGlyGlnLeu.....LysT 1066
3773 GACCGCAAAACCTTCGACAGCGGATCGCAACTCGCACGCGCTTGCC 3822
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1066 hGlnSerAsnArgTyrValLeuGlnLeuGlyGlyAspVal..... 1079
3823 CACGCGCGCGTTTTCGCGCATACGCGCATGCAAGTTTCATCGGCAT 3872
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1080 .....AlaGlnTyrSerGlnAsnGlySerAspArgTyrHisValGlyVal 1094
3873 CACGCGCGCGCGGCTTTAGC...ACGCGCAAGCTTCAGACGCGCATCG 3919
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1094 hMetAlaGlyTyrGlyAsnSerAspSerLysThrIleSerSerTrpThrG 1111
3920 GAGCGCAAAATCGCGCGCGCTGCTGCATTAC.....GGCATTCAGCA 3963
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1111 LTTTArgAlaAlaSerValAsnGlyTyrSerThrGlyLeuTyrAla 1127
3964 CGATACGCGCGGCTTTCGCGGATTCGCGCATCGACCGCATCGCGCG 4013
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1128 ThrTyrAla.....AspAspLysSerArgAsnGlyAl 1139
4014 AACGCGCTATTTCGTCAAAAAGCGATACCGGTACGAAAGCGTACATA 4063
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1139 a.....TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsnThr 1154
4064 TCGCACCGCGCGGCTTCATTCACCGCTACCGC..... 4098
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1154 aLysGlyAspAspLeuGlnSerLysSerTyrLysSerLysGlyPheThr 1170
4099 GCGGCGCATTAAGCAGATTAATTCATCAACCGCGCGCAACACATTCGC 4146
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1171 AlaSerLeuGlnAlaGlyTyrGlyHisLysLeuAlaGluPheAsnGlySe 1187
4147 .....ATCAGCGCTTATTGAGCGCTGCTCT 4171
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1187 rGlnGlyThrArgAsnGlnTyrTyrValGlnProGlnAlaGlnValThrT 1204
4172 ATACCGATGCC..... 4182
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1204 rPmetGlyValLysAlaAspLysHisArgGlnSerAsnGlyThrLeuVal 1220

```

```

4183 .....GCTTCGGGCAAAAGTCCGACACCGCGTCAATACCGCGTATT 4223
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1221 HisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysThrTr 1237
4224 GCGTCAG.....GATTTCGCAAAACCGC..... 4248
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1237 rLeuLysSerHisHisLysMetAspAspLysSerArgGluPheGlnP 1254
4249 .....AGTCCGAATGCGGCGGTAAACGCGCAATGCAAGGTTTCAG 4290
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1254 rPheValGluValAsnTrpLeuHisAsnSer.....LysAspPheSer 1268
4291 CTCTCCCTCCAGCGCTCCCGCGCGCAAGGCGCGCAATGCAAGCGCA 4340
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1269 ThrSerMetAspGlyValSerVal.....ThrGlnAspGlyAlaArgAs 1283
4341 CACGCGCGCATCAATTAAGC 4362
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1283 nIleAlaGluIleLysThrGly 1290
seq_name: p1r2:E85524

```

```

seq_documentation_block:
Probable beta-barrel outer membrane protein 20402 [imported] - Escherichia coli (strain
G:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001
C:Accession: E85524
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potlamousis, K.; Apod
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206531
A:Accession: E85524
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <STO>
A:Cross-references: GB:AE005174; NID:912513130; PIDN:AA654657.1; GSPDB:GN00145; UMG
A:Experimental source: strain O157:H7, substrain EDU933
A:Genetics:
A:Gene: 20402

```

```

alignment_scores:
Quality: 325.00 Length: 1474
Ratio: 0.509 Gaps: 72
Percent Similarity: 43.284 Percent Identity: 20.692

```

```

alignment_block:
US-09-303-518D-649 x E85524 ..

```

```

Align seg 1/1 to: E85524 from: 1 to: 1349

```

```

712 TCAGGTGGTGGCACAGTCACCTTAGTAGTGAATAAATTAACATAGCCC 761
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
118 SerGlyGlyGlyLeuAspIleThrGlyGlyValThrIleAsnGlyP 134
762 ATATGGTTTTTACCACAGGAGGCTCATTT.....GGCGACAGTG 802
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134 OleuAsnPheLeuLeuLysGlyThrGlyPheLeuAsnValSerAsnAlaG 151
803 GCTACCAAGATGTTTATGATGATGCCCAAAAGCAAAAGTGTTAAATTAAT 852
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151 LysSerGlyLeuTyrAlaAspAspLeuTyrGlnSer.....Asn 163
853 GGGGATTTGGCAAAAGGCGCAACCGCTATATAGCAAAAAGCAATGGCTTCA 902
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 SerGlyMetArgHisAspArgGlyTyrPheAsnValSerAsnGlyLys 180
903 GCTG..... 906
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
180 sIleHisValLysGlyThrSerArgLeuThrTyrLeuGlnGlyAsnValS 197
907 .....GTTCTGAAGATTCGTTCTATGATGAATAATCTTT 939

```

197 erGLyGLySerGIuValAsnSerGIuThrPhePheMetGIuValTyr 213
198
199 GCTGATTAAC.....GACGATAATAATGCGACGAAAAATCAATG 1030
200 |||||
201 rGIyGLyAsnGIuThrLeuSerValAsnAsnGIy...GIyGIuValAsnA 232
202
203 CCAACATGACACAT..... 1047
204
205 TCTCTGCTTAATAGATTAAAAACGACACCGTTCAATTG... 1086
206 |||||
207 LeuAlaValSerGIuGLyGIyLysIleSerAlaProThrIleSerLeuSe 265
208
209 TTTATGTTCTTATTCGACGACGACGACGACGACCGTTATCAATGCTG 1135
210 |||||
211 rThrsSerGIuLeuAlaLeuGIyAlaGIuGLySerAlaAlaLysA 282
212
213 CAGGTGTGTCTCAACGATTATCGACCCAGACTAATAATGAGAAATATT 1185
214 |||||
215 LaIaGLyLe.....IleAspAlaGIuLysIle 291
216
217 TCCTTTATTTGACGAAAGGACGAAATGATTAATTCACGACAC..... 1230
218
219 GIuPheValTyrAlaLysThrSerGIuLysIleThrLeuAsnHisTh 308
220
221 rAspLysAspAlaThrIleSerAlaAspIleValSerGIySerGIuLyl 325
222
223 ATCAATCAAGGTGCTGGAGATTAATTAATTCAGAGGAGATTTT 1272
224 |||||
225 euGIyTyrIleAsnAlaLeuAsnGIyThrTyrLeuThrGIyAsp... 340
226
227 ACGGTCTCCCTGAAAAATTAACGAACCTTGCAAGGCGCGGCTGTCTTAT 1322
228 |||||
229 AsnSerAlaPheSerGIyLys...ValLysIle 350
230
231 CAGTGAAGACAGTACCGTT.....ACTTGGAAG 1351
232
233 eGIuGIuAsnGIyAlaLeuGIyIleThrGIuAsnIleGIyThrAlaGIu 367
234
235 TAAACGCGCGCAACGACGCGCTGTCCAAATCGCAAGGACACCTG 1401
236 |||||
237 LeuAsn.....AsnArgGIyLysLeu 373
238
239 CACGTTTCAAGC.....AAAGGGGAAAAACA 1427
240 |||||
241 HisLeuValAlaAspSerMetThrPheAlaAsnLysIleSerGIyAs 390
242
243 AGGCTGATCAAGCTGGGCGGAGGTACATTTTGGATCAGAGGAG 1477
244
245 nGIyThrIleSerIleAspSerGIyThrValGIuLeu..... 402
246
247 ACGATAAAGCAAAAAACAACCTTACT...GAATCGGCTGTGTCAGC 1524
248
249 ThrGIyAsnAsnTyrAlaPheSerGIyTyrIleAspValAlaSer 417
250
251 GCGACG.....GCTACGCTGCAACT 1544
252
253 GIyAlaValAlaValIleSerGIuAspLysAsnIleGIyArgAlaGIuLe 434
254
255 GAATGCGATTAATCACTTC.....AACCCGACAAACTCTATT 1582
256
257 uAspValAspGIyLysLeuGIuIleAsnAlaAsnLysAspIrrValPhea 451
258
259 TCGGCTTTGCGGCGGACGT.....TTGGATTTAACGGCAT 1620
260

```
2314 .....GGGCTTGCACACTCAAC.....GGCAA 2336
      ||| |||||
720 AspleuArgSerGlyThrLeuLeuLeuAsnAspAsnValLeuGlyAs 736
2337 T.....CTTAGTCGAATGCGC 2353
      ||| |||||
736 nThrHisGluLeuArgLeuAlaIaGluThrGluLeuAspMetAsnGlyH 753
2354 ATACAGCTTTATACGTACGCCACACAGCCCAACCAAGCGCAACCTTAGC 2403
      ||| |||||
753 tSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeuLeuSer 769
2404 CTCGTGGCAATGCCCAAGCAACATTTATATCAAGCCCATTTAAAGCGCAA 2453
      ||| |||||
770 LeuAsnGlySerLeuThrValThrAsnGlyGlyThrSerThrGlySe 786
2454 CACATGGCTTCGGGCAATGCTTCATTTAATCTAAGCGCACCGCGCTAC 2503
      ||| |||||
786 rLeuThrGlySerGlyLeuLeuAsn.....IleG 796
2504 AAAAGCGCAGCTGTGACGCTTCCGGCAAGCTTAAGGCAACGTATAGCAT 2553
      ||| |||||
796 lngGlyThrLeuAspIleAlaGlyAsp.....Asn 806
2554 TCCGCACTCAACGTAATGTCTCCCTAGCCGATAGGCA.....GT 2594
      ||| |||||
807 SerAsnLeuThrAlaAsnValAlaIaAsnSerAlaAsnValLeuVa 823
2595 ATTCCATTTTGAAGAGCGCGCTTTACCGGACAAATCAGCGGCGGCAAG 2644
      ||| |||||
823 lSerHisAlaGlnGly..... 828
2645 ATACGGCATTTACCTTAAAGACAGCAATGAGACGCTGCCGTACGACAG 2694
      ||| |||||
828 ..... 828
2695 GAATTAGGCAATTTAAACCTTGAC...AACGCCACATTACACTCAATTC 2741
      ||| |||||
829 ...LeuGlySerAlaAsnValGluAsnAsnGlyThrLeuAlaLeuAsnAs 844
2742 CGCCTATGCCACGATGCGGAGG.....GGGCAACCGCGCA 2779
      ||| |||||
844 nSerAlaGluLysArgAlaAlaIaSerValAsnThrAlaLeuGlyAla 861
2780 GTGCGCAATGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTTA 2829
      ||| |||||
861 snLeuThrAsnAsnGly..... 866
2830 TTATCGGTTACACCGCCACTTCGGTGAATCCGCTTCAACAGCTGAC 2879
      ||| |||||
867 ...ThrLeuMetThrGlyMetSerGlyGlnAlaGlyAsnValLeuVa 882
2880 GGTAAAGCGCAATTTGACGCT...CAGGCAATTTCCGCTTTATGTGG 2926
      ||| |||||
882 lValLysGlyAsnGlyHisGlyAsnAsnGlyGlnLeuValMetAsnThrV 899
2927 AACTCTTGGGTCACCGCAGC.....GACAAATTAAGCTGGGGAAGT 2970
      ||| |||||
899 alLeuAsnGlyAspAspSerValThrAspLysLeuValValGlnGlyAs 915
2971 TCCGAAGCACTTACACCTTGGCGGTCAACATACCGGCAAGCAACCTGC 3020
      ||| |||||
916 ThrSerGlyThrThrAlaValThrValAsnAsnAlaGlyGlyThrGlyAl 932
3021 AAGCCTC.....GAACAATTAAGCGTATGGAAGAAAGCAACA 3061
      ||| |||||
932 aLysThrLeuAsnGlyIleGlnLeuIleHisValAspLysSerGlnG 949
3062 AACCGCTGCGCAAAACCTTAATTCACCCCTGCAAAACGAACAGCTGCAT 3111
      ||| |||||
949 lGlyLuphe.....ValGlnAlaGlyArgTrileVal 958

3112 GCCGGCGCGTGGCGTTACCACTCATCCGCAAAAGCGGAGTCCGCT 3161
      ||| |||||
959 AlaGlyAlaLysArgPyrThrLeuAlaArgGlyGlnGly..... 971
3162 GCATATCCGGTCAAGAACAGAGCTTCCGCAAACTCGGCAAGCGAG 3211
      ||| |||||
971 ..... 971
3212 AAGCCAAAAACAGCGCGGAAAAAGACAAACGCGCAAGCTTGACGCGTG 3261
      ||| |||||
972 .....AlaAsnSerGlyAsnTrpPyr 978
3262 ATTGCGCGCGCGCGCGATCGCTCGAAAAAGACAGAAAGCTTGCGCAAC 3311
      ||| |||||
979 LeuThrSerGlySerAspSerProGlu..... 987
3312 GCCCGCGCGAGGAGCGGGAATGTGGCATTTATGACAGGCGGAGAG 3361
      ||| |||||
987 ..... 987
3362 AAAAAAACGGGTGACAGCGGATAAAGACACCGCTTGCGCAACGCGC 3411
      ||| |||||
988 .....LeuGlnPro 990
3412 GAAGGGAAACCCGCGCGCTACACCGCTTCCCGCGCGCGCGCGC 3461
      ||| |||||
991 GluProAspProMetProAsnProGluProAsnProAsn..... 1003
3462 CGCGCGGATTTGCCCACTGCAACCCCAACCGGACCGGACCGGAGC 3511
      ||| |||||
1004 .....ProGluProAsnProAsnProThrProThrProGlyP 1016
3512 GCGACCTG.....ATCAGCGCTTTATGCCAATATGCGGTTG 3546
      ||| |||||
1016 roAspLeuAsnValAlaAsnAspLeuArgProGluAlaGlySerTryle 1032
3547 AGTGAATTTTCCGCGCACGCTCAACGAGCTTTGCGC.....GTACAGGA 3590
      ||| |||||
1033 AlaAsnLeuAlaIaIaIaIa...AsnThrMetThrThrArgLeuHisGly 1048
3591 CGAATTA.....GACCGGCTATTTGCCGAAGACCGCC 3622
      ||| |||||
1048 uArgLeuGlyAsnThrTrpThrTrpAspMetValThrGlyGlnGlnLysG 1065
3623 GCAACGCGCGTTGGCAAGCGGCAATCGGCGGACACCAAACTACCTTCG 3672
      ||| |||||
1065 lThrThrMetTrpMetArg.....HisGlyGlyGlyHisAsnLysTrp 1079
3673 CAAGATTTCCGCGCTACCGCCAAACCAACGACCTGCGCAAAATCGGTAT 3722
      ||| |||||
1080 ArgAsp..... 1081
3723 GCAGAAAAACCTCGCGACGCGCGCGCTGCGCATCTGTTTCGCAACAC 3772
      ||| |||||
1082 .....GlySerGlyGlnLeu.....LysT 1088
3773 GGACGAAACACCTTCGACAGCGCATGCGCAACTGCGGACGCGCTTGGC 3822
      ||| |||||
1088 hrGlnSerAsnArgTrpValLeuGlnLeuGlyGlyAspVal..... 1101
3823 CACGCGCGCGTTTGGGCAATACGCGATCGACAGGTTCTACATCGGCAT 3872
      ||| |||||
1102 .....AlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValGlyVa 1116
3873 CAGCGCGGCGCGGCTTTAGC...AGCGGACCGCTTTCAGAGCGCATCG 3919
      ||| |||||
1116 lMetAlaGlyTrpGlyAsnSerAspSerLysThrIleSerSerArgThrC 1133
3920 GAGCAAAATCCGCGCGCGCGCTGCTGATAC.....GGCATTCAGGCA 3963
      ||| |||||
1133 lYTrArgAlaLysAlaSerValAsnGlyTrpSerThrGlyLeuGlyAla 1149
3964 CGATACGCGCGGCTTTCGCGGATTCGGCATCGCAACCGCATCGCGCGC 4013
```



```

242  GlnuInlleValIgluValAlaAsnGluThrThrIleAsnAspGly 258
      ::::: |||::: ::||
2316 GCTTGCACCTCAACGGCAATCTTAGTCAATGGCGAT.....A 2356
      ||| ||| ::||| |||:::
258  YIleGlnThr.....ValSerAlaAsnGlyGluAlaIleLysr 271
      ::::: |||:::
2357 CACGTTATACAGTCAGCACACACCGCAAC..... 2397
      |||::: |||::: ||| ||| |||
271  hrlsIleAsnGluGlyGlyThrLeuThrValAsnAspAsnGlyLysAla 287
      ::::: |||:::
2398 CTTAGGCTGTGGCAATGCGCAAGCAACATTAAATCAAGCCACTTAA 2447
      ::::: |||::: |||::: |||::: |||:::
288  ThrAspIleValGlnAsnSerGlyAlaIleGlnThrSerThrAlas 304
      ::::: |||:::
2448 CGGCAACACATCGGCTTGGGCAAT.....GCTCATTTAACTAA 2488
      |||::: |||::: |||::: |||::: |||:::
304  nGly...IleGluIleSerGlyThrIleGlnThrGlyThrPheSerIleS 320
      ::::: |||::: |||::: |||::: |||::: |||:::
2489 GCGACACACCGCTACAAAAGCGAGCTGACCTTCCGGCAACGCTAAG 2538
      |||::: |||::: ||| ||| |||
320  eTcIyAsnLeuAlaThrAsnMetLeuGluAsnGlyLysAsnLeu 336
      ::::: |||::: |||
2539 GCAAAAGCTAAGCATCCGACATCAACGGTAATGTCCTCCCTAGCCGAT 2588
      ::::: |||::: |||::: |||::: |||::: |||:::
337  ValLeuAlaGlyThrGlnAlaArgAspSerThrValGlyLysGlyGly.. 352
      ::::: |||::: |||::: |||::: |||::: |||:::
2589 GCGAGTATTCCATTGTAAGACGCGCTTACCGGACAAATCAGCGGCG 2638
      |||::: |||::: |||::: |||::: |||::: |||:::
353  .AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsnSerGlyG 369
      ::::: |||::: |||::: |||::: |||::: |||:::
2639 GCAGGATCGGCGATTACACTTAAAGACAGCGAATGACGCTGCCGTCA 2688
      |||::: |||::: |||::: |||::: |||::: |||:::
369  LyGlnThrLeuGlyArgSerLysAspGluPheGlnAlaLeuAlaArg 385
      ::::: |||::: |||::: |||::: |||::: |||:::
2689 GCGACGGAATTA..... 2700
      ::::: |||::: |||::: |||::: |||::: |||:::
386  AlaGluAspLeuGlnValAlaGlyLysThrAlaIleValIYrAlaGlyTh 402
      ::::: |||::: |||::: |||::: |||::: |||:::
2701  .....GCAATTAAACCTT... 2715
      ::::: |||::: |||::: |||::: |||::: |||:::
402  rLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSerLeuMet 419
      ::::: |||::: |||::: |||::: |||::: |||:::
2716  .....GACAAAGCCAGC...ATTACATCAATCCGCGCTATCC... 2751
      |||::: |||::: |||::: |||::: |||::: |||:::
419  hrProAlaAspAsnValThrProValLysLeuGlyAlaValArgIle 435
      ::::: |||::: |||::: |||::: |||::: |||:::
2752 CACGATCGGCGAGGCGCAACCGCAGTGCAGACAGATCGCGCGCG 2801
      |||::: |||::: |||::: |||::: |||::: |||:::
436  ThrAspSerAlaThrLeuThrLeuGlyAsnGlyValAspThrThrLeuAl 452
      ::::: |||::: |||::: |||::: |||::: |||:::
2802 CGGTCGCGCGGTCGCGCGCT... 2823
      ::::: |||::: |||::: |||::: |||::: |||:::
452  AspLeuThrAlaIleSerArgGlySerValITrPLeuAsnSerAsnAsn 469
      ::::: |||::: |||::: |||::: |||::: |||:::
2824  .....TCCCTATTA... 2832
      ::::: |||::: |||::: |||::: |||::: |||:::
469  eTcYsAlaGlyThrSerAsnGlyLysValArgValAsnSerLeuLeu 485
      ::::: |||::: |||::: |||::: |||::: |||:::
2833  .....TCCGTTACACCGCGCACTTCCGCT 2855
      ::::: |||::: |||::: |||::: |||::: |||:::
486  AsnAspGlyAspValITyTLeuSerAlaGlnThrAlaIleProAlaThrTh 502
      ::::: |||::: |||::: |||::: |||::: |||:::
2856 AGAATCCGCTTAAACAGCGTACGAGTAAGCGAAATGACAGCGTACG 2905
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
502  rAsnGlyIleTyrAsnThrLeuThrAsn...GluLeuSerGlySerG 518
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
2906 GAAATCCGCTTATGTCGAACTTCTGCTACGCGACGACGAAATTC 2955
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
518  LysnRheITyLeuIleThrAsnValAlaGlySerArgGlyAspGlnLeu 534
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
2956 AAGTGCAGAAAGTCCGAAGCACTTACACCTTGGCGGTCAACATAC 3005
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::

```

```

535  ValValAsnAsnAlaThrGlyAsnRheLysIlePheValGlnAspTh 551
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3006 CGGCAACGACCTGCAACAGCTGCAACATTCAGCGTACTGACAGAAAG 3055
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
551  rGlyValSerProGlnIleAspAlaMetThrLeuVal..... 564
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3056 ACAACAAACCGCTGCGCAAAACCTTAATTCACCTGCAAAAGAA... 3102
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
565  ....LysThrGlyGlyLysAspAlaSerPheThrLeuGlyAsnThrGly 579
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3103  ...CACGTCGATGCGCGCGGGGTACCACTCATCCGCAAGACG 3149
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
580  GlyPheValAspLeuGlyThrTyrGluTyr...ValLeuLysSerAspG 595
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3150 CGAGTCCGCTGCTAATCGGTCAAGAACAGAGCTTCCGACAAAC 3199
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
595  Y.AsnSer..... 597
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3200 TCGGCAACGAGAAAGCAAAACAGCGGCAAAAGACAGCGCAAGC 3249
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
598  ..... 600
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3250 CTTGACGCGCTGATTCGCGCGCGCGGAGTCCGCAAAAGACAGAAAG 3299
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
600  nLeuThrAsnAspValLysProAsnProAspPro..... 611
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3300 CGTTGCCGACCGCGCGCGGAGCGAGCGGGGAAATGTCGGCATTTATGC 3349
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
612  ..IleProAsn..... 614
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3350 AGGCGAGAGAAAGAAACCGGGTGCAGGCGGATTAAGACACCGCTTG 3399
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
614  ..... 614
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3400 GCGAACAACGCGGAGGAAACCGCGCGCTACACACCGCTTCCCGCG 3449
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
615  .....ProLysProAspProLysProAspProLysProLysPro 626
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3450 CGCCGCGCGCGCGCGCGGATTTGCGCACTGCAACCCCAACCGCAGC 3499
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
626  sPro..AspProLysProAsp...ProThrProAspProThrProThrP 641
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3500 CCNACCGCAGCGGAGCTGATCAGCGCTTATCCATACGCTTGAGT 3549
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
641  roValProGluLysArgIleThrPro.....SerThrAlaIleValLeu 655
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3550 GAATTTCCCGCAGCGTCAACACGCTTTC.....GCGCT 3584
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
656  AsnMetAlaIleThrLeuProLeuValPheAspAlaGluLeuAsnSerI 672
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3585 ACAGAGCAATTAGACCGCGTATTTGCCGAAGACCGCGCAACGCGGTT 3634
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
672  eArgGluArgLeuAsnIleMetLysAlaSerProIleAsnAsnValIT 689
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3635 GGAACAACGCGCATCGGAGCACCAACACTACGCTTCCGAGATTTCCG 3684
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
689  rp...GlyAlaThrTyrAsnThrArgAsnValIThrThrAspAlaGly 704
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3685 GCCTACCGGCAACACGAGCTGCGCAATCGGATGACAAAAACCT 3734
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
705  AlaGlyPheGluGlnThr.....LeuThrGlyMetThrValGlyI 718
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3735  C.....GCGACGCGCGCGCGCGGACATCCGCT 3760
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
718  eAsp***ArgAsnAspIleProGluGlyIleThrThrLeuGlyAlaRhe 735
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3761 TTTGCGACAAACGAGCGAACAACCTTCCGACGAGCGC..... 3798
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
735  eTcIyTyrSerIleSerIleIleGlyPheAspArgGlyGlyIleGlySer 751
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
3799  .....ATCGCAACTCGGACGCGCTTCCGCAACGCGCGCGCT 3833
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
752  ValGlySerTyrSerLeuGlyGlyTyrAlaSerITrGluHisGluSerG 768
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::

```



```
1974 GTCGCAAAAAGAGGGCATTCCT...CGGGGGAATGCTGTGGACAACG 2020
    |||
339 pThr.....ValProIleThrGlyThr.....AspHis 349
2021 ACTGGATCAACCGCAGCATTTAAAGCGGA.....AAC 2052
    |||
349 erpThrPheThrProTyrAlaIleArgThrAspArgIleGlyIleAsn 365
2053 TTCCAATTAAGAGCGGACGACGGCGGTTCCTCCGCAATGTTGGCAACGT 2102
    |||
366 TyrPheAsnGlyGlyGlyValValGluSerThrThrSerGln.. 381
2103 GAAAGGCGATTGGCATTGAGCAATCAAGCCCAAGCAGTTTGGGTGCG 2152
    |||
382 .....SerLeuSerGlnSerLysSerLeuSerValSerIAs 394
2153 CACGCGATCAAGCCACACAACTGTACAGCTTGAGCTGGACGGGTCTG 2202
    |||
394 erGlnSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer 410
2203 ACAATTTGTGCAAAAAACCATTAACCGCAATAAAGTATGCTTCAT 2252
    |||
411 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer 426
2253 GACTAAGACCGACATCAGCGGCAATGTCATTTGCCGATCAGCTCAT 2302
    |||
427 .ThrSerAlaSerValSerAlaSerThrSerAlaSerAlaSer 443
2303 TAAATTCACAGGGCTTGCCACACTGACGCAATCTTACGCAATGCG 2352
    |||
443 hrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerThr 459
2353 GATACACGTTATACAGTCACGCCAACGCGCCAAAGCGCAACCTTAG 2402
    |||
460 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerThrSe 476
2403 CCGCGGGGCAATGCCCAAGCAGCATTTAATCAGCCACATTAAACGCG 2452
    |||
476 rAlaSerAlaSerAlaSerThrSer.....AlaSerGluSerIAs 490
2453 ACACATCGGCTTCGGCAATGCTTCATTTAATCTAAGCAGACCGCGCG 2502
    |||
490 erThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerIAs 506
2503 CAAAGCGCAGCTGACGCTTCGGCAACGCTAAGCAGCAACGTAAGCCA 2552
    |||
507 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerThr 523
2553 TTCGCGACACAGGTAATGTCCTCCCTAGCCGATAAGCAGATTCATT 2602
    |||
523 rSerAla...SerGlySerAlaSerThrSerThrSerAlaSerAlaSer 539
2603 TTGAAGACGCGCGCTTTACCGGACAATATCGCGGCGCAAGGATACG 2652
    |||
539 hrSerAlaSerAla...SerAlaSerThrSerAlaSerAlaSer 554
2653 TTAACTTAAGACAGCAATGACGCTGCCCTAGCAGCAGGAATTAGG 2702
    |||
555 lIeSerAlaSerGluSerAlaSerThrSerAlaSerGluSerThr 571
2703 CAATTTAACTTGAACAGCCACCATTAACCTAATCCGCTATCCGC 2752
    |||
571 rSerThrSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThr 588
2753 ACAGTGGCGGAGGGCGCAACCGGCACTGCGACAGATGCGCGCGCGC 2802
    |||
588 erAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThr 604
2803 CGTTGGCGCGCGCTTCGCGCTTATTCCTTATCCGTTACCGCCCACTC 2852
    |||
605 AlaSerAlaSerAlaSerThrSerAlaSerAlaSerThrSerAlaSer 621

2853 GGTAGATCCCTGTTTACACGCTGACGGTTAAAGCGCAATTAAGCGTC 2902
    |||
621 uSerAlaSer..ThrSerAla.....SerAlaSerAlaSerThr 633
2903 AGGAAACATTCGCTTTATGTGCGAACCTCTTGCGCTACCGCAGCCACAA 2952
    |||
634 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThr 650
2953 TTGAAGCTGGCGGAAAGTTCCGAGAGCAGCTTACCTTGCGGTCAACA 3002
    |||
650 rAlaSerAlaSerAlaSerThrSerAla.....SerValS 662
3003 TACCGCAACGAAACCTGCAAGCTTGCAACAATTGACGATGAGGAAGAA 3052
    |||
662 erAlaSerThrSerAlaSerAlaSerAlaSer.....Thr 673
3053 AACACACAAACCGCTGTCGAAACCTTAATTCCTGCAAAACGAA 3102
    |||
674 SerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThr 690
3103 CACGTGATCGCGCGCGCGGTGCGCTTACCAACTATCCGCAAGCGCA 3152
    |||
690 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThr 707
3153 GTTCGCGCTGCATTAATCCGTCAAAGAACAGCTTCCGCAAACTCG 3202
    |||
707 lSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 723
3203 GCAAGGCAAGACCAAAACAGCGGCAAAAGACAGCGCAACGCTT 3252
    |||
724 Ser..AlaSerAlaSerThrSerAlaSerGluSerAlaSerThrSerAla 739
3253 GACGGCGCTGATGCGCGCGCGCGCATCCGTCGAAAGACAGAAAGCGT 3302
    |||
740 SerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 756
3303 TGCCGAACCGCGCGCGGAGCGGCGGAAATGTGCGCATTTGCGAG 3352
    |||
756 rAlaSerAlaSerThrSerAlaSerAla..... 765
3353 CGGAGGAAGAGAAAAACGGGTGCGAGCGGATTAAGACCGCGCTTGGCG 3402
    |||
766 .....SerAlaSerThrSerAlaSerAlaSerAlaSerIleSer 778
3403 AAACGCGCGAAAGCGGAAACCGCGCGGTACACCGCTTCCCGCGCGC 3452
    |||
779 AlasSerGluSerAlaSerThrSerAlaSerAlaSerAla..SerThrSerAla 795
3453 CCGCGCGCGCGCGCGGATTTGCGCAACTGCAACCCCAACCGCGCGCC 3502
    |||
795 lAsSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 811
3503 AACCGCAGCGGACCTGATCAAGCGTTATAGCAATAGCGTTGAGTAA 3552
    |||
812 SerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 828
3553 TTTCCCGCAGCTCAACAGCGCTTTGCGCGTACAGACCAATTAAGCCG 3602
    |||
828 rAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSer 845
3603 CGTATTTGCGGAGACCGCGCAACGCGCTTTGAGCAAGCGGATCCGG 3652
    |||
845 lAsSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAla 861
3653 ACACCAAAACACTACCGTTCCGCAAGATTTCCGCGCTTACCGCCACAAACC 3702
    |||
862 SerAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAla 878
3703 GACCTGGCCCAAACTCGGTATGCAAAAAACCTGCGACGCGGCGCTGCG 3752
    |||
878 rAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSer 895
3753 CATCTGTTTTCGACACCGGACGCAAAACACTTTCGACGAGCGGATCG 3802
```

```

|||||
895 Laser.....ThSerAlaSerAlaSerThSerAlaSer 908
3803 GCAACGCGGACGCGCTTGGCCAGCGCCGCTTTTCGGCAATACGGCATC 3852
|||||
909 AlaSerAlaSerThSerAlaSerAlaSerThSerAlaSerAlaSer 925
3853 GACAGGTTTACATCGGCATCAGCGCGCGGCTTTAGCAGCGGACG 3902
|||||
925 aSerAlaSerThSerAlaSerAlaSerThSerAlaSerAlaSer 942
3903 CCTTCAGCGCGCATCGGCAAAATCCGCGCGCTGCTGATTCATACG 3952
942 erThSerAlaSerAlaSerThSerAlaSerAlaSerAlaSer 958
3953 GCATTACGACGATACCGCGCGCTTTCGGCGGATTCGGCATCGAACG 4002
959 ThSerAlaSerAlaSerAlaSerThSerAlaSerAlaSerThSer 975
4003 CACATCGCGCGACGCGCTATTTCGTCMAAAACGGATTAACGCTACGA 4052
975 rSerAlaSerAlaSerAlaSerThSerAlaSerAlaSerAlaSerTh 992
4053 AACGTCATATCGCACCCCGCGCTTTCGATTCACCGCTACCGCGCG 4102
992 erAlaSerAlaSerAlaSerThSerAlaSerAlaSerAlaSerThSer 1008
4103 GCATTACGACGATTTATTCATTCAACCGCGCGCACATTCATTCACG 4152
1009 AlaSerAlaSerAlaSerThSerAlaSerAlaSerAlaSerAlaSer 1024
4153 CCTTATTGAGCGCTGCTTATACGATGCGCTTTCGGCGCAAGTCCGAC 4202
1025 AlaSerAlaSerAlaSerThSerAlaSerAlaSerAlaSerThSerAla 1041
4203 AGCGCTCATATACCGCGCTATTGCTCAGAGATTTCGGCAAAACCGCGAC 4252
1041 aSerAlaSerAlaSerThSerAlaSerAlaSerAlaSerThSerAlaSer 1058
4253 CGGATGCGCGCTTAACCGCGCAATCAAGGTTTCAGCGCTGCGCTCCAC 4302
1058 erGlSerAlaSerThSerAlaSerAlaSerAlaSerThSerAlaSer 1074
4303 GCTGCGCGCGCCAAAGCGCGCACTGGAGCGCAACACAGCGCG 4347
1075 GlSerAlaSerThSerAlaSerAlaSerAlaSerThSerAlaSer 1089

seq_name: pir2:A47675

seq_documentation_block:
68k outer membrane protein P.68 pertactin - Bordetella bronchiseptica
C:Species: Bordetella bronchiseptica
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A47675
R:Li, J.; Fairweather, N.F.; Novotny, P.; Dougan, G.; Charles, I.G.
J. Gen. Microbiol. 138, 1697-1705, 1992
A:Title: Cloning, nucleotide sequence and heterologous expression of the protective oute
A:Reference number: A47675; MUID:92407514
A:Contents: CN7531
A:Accession: A47675
A:Status: Preliminary
A:Molecule type: nucleic acid
A:Residues: 1-911 <L11>
A:Cross-references: GB:X54815; GB:S46416; NID:g39396; PID:CAA38584.1; PID:g39397
A>Note: sequence extracted from NCBI backbone (NCBIN:113318, NCBI:P.113319)

alignment_scores:
Quality: 295.50 Length: 1079
Ratio: 0.612 Gaps: 50
Percent Similarity: 44.764 Percent Identity: 21.965

alignment_block:

```

```

US-09-303-518D-649 x A47675 ..
Align seg 1/1 to: A47675 from: 1 to: 911

1417 GGGGAAACCAAGGCTGCATCAGCGCGGACGCGTACAGCATTTTGA 1466
45 GlyIuArgIuInHIsIle..... 53
1467 TCAGAGGCGACGATTAAGGCAAAACGCTTTAGTGAATCGGCT 1516
54 ..GlnSerAspGlyAlaGlyAlaArgThAlaThrGlyThrIleL 70
1517 TGGTACGCGCGAGGCGTACGCGTCACTGAATCCGATATCATTCAC 1566
70 yValSerGlyArgIuInHIsIleGlyAlaLleuLeuGluAsn..... 83
1567 CCCGCAAACTGATTTTCGCTTTCGGCGGCGGCGCTTGGATTTAAACG 1616
84 ..ProAlaIleGluLeuArgPheGlnHIsIleGlySerValThSerG 99
1617 GCATTGCTTTCGCTTTCACCGCTATTCAAATACGATGAAGGCGGATGA 1666
99 Y.....GlnLeuPheAspGlyVal.... 106
1667 TTGTACACCAATCAAGCAAAAGATTCACCGCTTACCTTACAGCAAT 1716
107 .....ArgArgPheLeuGlyThrValThValLysAlaGly 118
1717 AAT.....GATATGCTACACCGCAATTAACACGCTTGGATAG 1757
119 LysLeuValAlaAspHisAlaThrIleuAlaAsn..... 129
1758 CAAAAAAGAAATTCCTACACGCTTGGCTTTCGGCAAAAGATACGACA 1807
130 .....ValSerAspThrArg 135
1808 AATGCAAGCGGCGGCTCAACCTTGT.....TACGACCGCGC 1845
135 spAspAspGlyIleAlaLeuTyValAlaGlyIuInHIsIleAlaSer 151
1846 GCAGAGACCGCACCGCTGCTTTCGGCGGCAAAATTA.....AA 1889
152 IleAlaAspSerThrIleuInHIsIleGlyAlaArgValGluArgI 168
1890 CGGCACATATCGCAAAACGCGCAACGCTTTTCAGCGG..... 1932
168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyLeuHisI 185
1933 .....AGACCAACACCGCGCTACATCATTTAAACGACAT 1971
185 IeGlyThrIleuGlnProLeu..... 191
1972 TGGTGCAGAAAGAGGCGATTCCTCGCGGGAATGCTGGGACACACA 2021
192 .....GlnProGlnAspLeuProProSerArgValAlaLeuGlySp 206
2022 CTGATATCAAC.....CGACAT 2038
206 rSerValThAlaValProAlaSerGlyAlaProAlaAlaValSerAl 223
2039 TTAAGCGGAAACTTCCAAATTAAGCGGACGCGGCTGTTCCCGC 2088
223 heGlyAlaAsnGluLeuThrValAspGlyHisIleThrGlyArg 239
2089 AATGTTGCCAAGATGAAGGCGATTTGGCATTCAGCAATACGCGCAAC 2138
240 AlaAlaGly.....ValAlaAl 245
2139 AGTTTTCGCTGCGACCGCATCAAGCGCACACATCTGTACAGCTTCG 2188
245 ametAspGlyAlaIleValHIsIleuGlnArgAlaThrIleArgArgG 262
2189 AC..... 2190

```

```

262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyPhe 278
2191 .....TGGAGCGGCTCTCAACAATTGTCTCGAAAAAAC 2222
279 GlyProLeuLeuAspGlyTTPryGlyVal..... 288
2223 CATTACGCGAGATAAGTGAATGCTTCATTGACTAAGACCGACATCAGC 2271
289 .....AspValSerA 292
2272 ..GGCAATGTCATCTTGCCGATCAGCTCATTTAAATCTCAGAGGCTT 2319
292 spSerThrValAspLeuAlaGln.....SerIleValGluAla 304
2320 GCCACACTCAAGCAATCTTAGTGAANAAGGAGTACACGTTATTCAGT 2369
305 ProGlnLeuGlyAlaIleAlaIleAlaGlyAlaGlyAlaIleAlaValThrVal 321
2370 CAGCCACAAAC.....GCCACCCAAAACGCGACACTTAGCTCTGGGCA 2413
321 IserGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGly 338
2414 ATGCCCAACGACATTTAAT.....CAAGCCACATTA 2445
338 IyGlyAlaIleArgPheProProAlaSerProLeuSerIleThrLeu 354
2446 AAGCGCAACATCGGCTTCGGGCAATGCT..TCATTTAACTAAGCA 2492
355 GlnAlaGlyAlaIleAlaGlnGlyAlaLeuLeuIleValLeuIle 371
2493 CCACGCGGTACAAAACGCGAGTTCGAGCTTCGCGCAACGCTAAGCA 2542
371 OGluProValIys.....LeuThrLeuAlaGlyGlyAlaGlnGly 384
2543 ACGTAGCATTCGCGACTCAACGGTAATGTCTCCCTACCGCATAAGCA 2592
385 .....GlnGlyAspIleValAlaThr..... 391
2593 GATTTCCATTTGAAAAGCGCGCTTACCGGCAAAATCAGCGGGGCA 2642
392 .....GluLeuProIleProGlyAlaSerSerGlyProIle 404
2643 GGATACGGATACACTTAAAGACAGCGAATGACGCTGCGCGAGCA 2692
404 uAspAlaIleLeu..AlaSerGlnAlaIleArgTrpPhe..GlyAlaThrA 419
2693 CGGAATTAGCAATTTAAACCTTGACACGCCACCATTAACCTCAATTC 2742
419 rgAlaValAlaSerLeuSerIleAspAsnAlaThrTrpValMet..... 433
2743 GCCTATCGCAGATCGCGAGGGGCGCAACCGGCACTGGCAGATGC 2792
434 .....ThrAsp..... 435
2793 GCCGGCGCGCGTTCGCGGTCGCGCGCTTCCTATTATCCGTTAC 2842
436 .....AsnSerAsnValGlyAlaLeuArgLeuAla 446
2843 CGCCAACTCGGTAGAA.....TCCGTTTCAAC 2871
446 erAspGlySerValAspPheGlnGlnProAlaGluAlaGlyArgPheIys 462
2872 ACGTCAGCGTAACGCAATTTAGCGGTAGGGAACATTCGCTTTAT 2921
463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478
2922 GTCCGAATCTTTCGCTACCGCAGCAAAATTTGAAGCTGGCGCAAGTT 2971
478 nValPheAlaAspLeuGlyLeuSerAspIysLeuValValMetCArgAspA 495
2972 CCGAAGCACTTACACTTGCGGCTCAACAATACGGCAAGCAACTGCA 3021
.....
495 IAserGlyGlnHisArgLeuLeuValArgAsnSerGlySerGluProAla 511
3022 ACCCTGCAACAATTTGACGTAGTGAAGAAAAGAACAAACAAACGCTGTC 3071
512 SerGlyAsnThrMetLeuLeuValGln.....ThrProArgIle 524
3072 CCAAAAACCTTAATTTACCTTGCAAAACGAA.....CAGCTGATGCCG 3115
524 ySerAlaIleAlaThrPheThrLeuAlaIleAsnIysAspGlyLysValAspIleG 541
3116 GCGCGGCGGTATACCAACTCATCCGCAAAAGACGGGAGTTCGCGCTGAT 3165
541 IyThrTyrArgTyrArgLeu..... 547
3166 AATCGGTCAAAAGAACAGAGCTTCGCAAAACTCGCAAGCGAGAAC 3215
547 ..... 547
3216 CAAAAAACGCGGCAAAAAGAACAGCGCAAAAGCCTTGACGCGGTATG 3265
548 .....AlaIleAsnGlyAsnGlyGlnTrp.....SerLeuValG 559
3266 CGCGCGGCGCGGATGCGGTGCAAAAAGAACAGAACGCTTGCCGAACGGCC 3315
559 IyAlaIleAlaProProAlaPro.....LysPr 568
3316 CGCGAGCGAGCGGGGAAATGTCGCGCATTTATGACGCGGAGAGAGAA 3365
568 OAlaProGlnProGlyPro..... 574
3366 AAAAGCGGTGACGCGGATTAAGACACCGCCTTGCGAAGACGCGAAG 3415
574 ..... 574
3416 CGAAACCGCGCGGCTACACACGCTTCCCGCGCGCGCGCGCGCG 3465
575 ..GlnProGlyProGlnProProGlnProProGlnProProGlnPro 590
3466 CGGATTTGCGCGCAACTGCAACCCCAACCGCAGCGCCACCGAGCGGA 3515
590 roGlnArgGlnProGluAla..ProAlaProGlnProProAlaGlyArgIle 606
3516 CCTATACGCGGTATAGCAATAGCGGTGTGAGTGAATTTCCGCCACGC 3565
606 uLeu...SerAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 622
3566 TCAACAGCGTT.....TTCGCGGTACAGGAGCAATTTAGCCGATTT 3609
622 euAlaSerThrLeuTrpTyrAlaGlnSerAsnAlaLeuSerIysArgLeu 638
3610 GCCGAACACCGCGCGCAAC.....GCCGTTTGACAAAGCGCAT 3647
639 GlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaIleTrpGlyArgIle 655
3648 CCGGAGACCAACCACTACCGCTTCGCAAGATTTCCGCGCTACCGCGCAAC 3697
655 euAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArgPheAspGln 672
3698 AAACCGACCTCGCGCAAAATCGGTATCGCAAGAAAACCTC..... 3735
672 ysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAlaValAlaGly 688
3736 GCGAGCGGCGCGGTGCGCATCTGTTTGGCAACCGGAGCGGAACAC 3785
689 GlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgIleAspArgIle 705
3786 CTTT...GACGAGCGCATCGCAACTCGCAGCGCTTGCCGAGCGCG 3832
705 yPheThrGlyAspGlyGlyGlyHisThrAspSerValHisValGlyGlyT 722
3833 TTTTGGCAATACGGCATCGACAGTTCATATCGGCTACAGCGCGGCG 3882
722 yAlaIleThrTyrIleAlaAsnSerGlyPheThrLeuAspAlaThrLeuArg 738

```

```

3883 GCGGCTTTTAGCAGCGCAGCCTT.....TCAGACGC..... 3915
||||| :||||| :|||||
739 AlsterArgLeuGlusnAspPheLysValAlaGlySerAspGlyTyrAl 755
3916 .ATCGAGGCAAAATCCCGCCGCGCTGCTGCTTACGGCATTCAGGCA 3964
::: ||||| :||| :||||| :|||
755 aValLysGlyLysTyrArgThr.....HisGlyValGlyAlas 768
3965 GATACCGCGCCGCT.....TTCGCGCATTCGGCATCGAA 3999
:::||||| :||||| :|||||
768 erLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTyrPheLeuGlu 784
4000 CCGCAGCATTCGCGCAACGCGCTATTGCTCCAAACCGCATTCACGCTA 4049
||||| :||| :||| :||||| :|||||
785 ProGlnAlaGluLeuAlaValPheArgValGlyGlySerTyrArgAl 801
4050 CGAAAC...GTCAATATGCGCAACCGCGCTTGCATTCACCGCTACC 4096
||| :||| :||| :|||
801 aAlaAsnGlyLeuArgValArgAspGlyGlySerSerValLeuGlyA 818
4097 GCGCGGCGCATTAAGCAGATTATTCATCAACCGCGCAACACATTC 4146
||| :||||| :||| :||| :|||
818 rglLeuGlyLeuGluValGlyLysArgGlyLeuAlaGlyGlyArgGln 834
4147 ATCAGCGCTTATTCAGCCTGCTCTATACGATGCCGCGCAAGT 4196
::: ||||| :||| :||| :|||
835 ValGlnProGlyTyrLeuAlaSerValLeuGlnLeuPheAspGlyAlaG 851
4197 CCGAAGCAGCGCTACATACCGCGCTATTGCTCGCATTCGCGCAACCC 4246
:::||||| :||| :||| :|||
851 yThValArgThrAsnGlyLeuAlaHisArgThrGluLeuArgGlyThra 868
4247 GCACTGCGGGAATGGCGGTAAACCGCGCAATCAAGTTTCAGCCTTC 4296
||| ||||| ||| :||| :|||
868 rg...AlaGluLeuGly.....LeuGly 874
4297 CTCGACGCTGCGCGCCCAAGCGCGCACTGGAAGCG.....CAACA 4340
:::||||| :||||| :||| :|||
875 MetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGly 891
4341 CAGCGCGGCGCATCAATTAAGCTACCGCTG 4371
||| ||| ||||| :|||
891 rSerLysGlyProLysLeuAlaMetProTyr 901

seq_name: p1r2:S15204

seq_documentation_block:
pertactin - Bordetella parapertussis
N:Alternate names: outer membrane protein p70
C:Species: Bordetella parapertussis
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 08-Oct-1999
C:Accession: S15204; S14659
R:Li, L.J.; Dougan, G.; Novotny, P.; Charles, I.G.
Mol. Microbiol. 5, 409-417, 1991
A:Title: P.70 pertactin, an outer membrane protein from Bordetella parapertussis: clonin
A:Reference number: S15204; MIMD:91251771
A:Accession: S15204
A:Molecule type: DNA
A:Residues: 1-922 <LIL>
A:Cross-references: EMBL:X54547; NID:g939761; PIDD:CA38419.1; PID:g939762
C:Genetics:
A:Gene: prn
C:Keywords: membrane protein

```

```

alignment_scores:
  Quality: 295.00      Length: 1083
  Ratio: 0.602        Gaps: 52
  Percent Similarity: 45.245      Percent Identity: 22.253

```

alignment_block:
US-09-303-518D-649 x S15204

```

Align seg 1/1 to: S15204 from: 1 to: 922
1417 GGGGAAACCAAGGCTGATCAGCGTGGCGACGCTACAGCATTTTGA 1466
||||| :||| :||| :|||
45 GlyIuArgGlnHisGlyLeuHisLe..... 53
1467 TCAGCAGCAGCATTAAGCAAAACAGCCTTAGTGAATGCGAT 1516
:::||||| :||| :||| :|||
54 LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrLe 70
1517 TCGTCAGCGGAGGCGGTACGTTGCAACTGAATGCCATATCACTTAAC 1566
||||| :||| :||| :|||
70 yValSerGlyArgGlnAlaGlnGlyValLeuLeuGlnAsn..... 83
1567 CCCGCAACACTTATTCGCTTGGCGGGGAGAGCTTGAATTAACG 1616
::: ||| :||| :||| :|||
84 ..ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerG 99
1617 GCATTCGCTTCTTCACCGCTATTCAAAATACGATAGAGCGCGATGA 1666
||| :||| :||| :|||
99 y.....GlnLeuPheAspGlyVal.... 106
1667 TTGTCACACCATCAAGCAAAAGATCCAGCGTACCATTAAGCAAT 1716
::: ||||| :||| :||| :|||
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGly 118
1717 AAA.....GATATTGCTACACCGCAATTAACACAGCTTGATAG 1757
||| ||||| :||| :|||
119 LysLeuValAlaAspHisAlaThrLeuAlaAsn..... 129
1758 CAAAAAGAAATTCCTACACGCTTGCTTGGCGCAAAATACGACCA 1807
:::||||| :||| :|||
130 .....ValSerAspThrArgA 135
1808 AAACGAGCGCGCGCTCAACCTGTT.....TACAGCGCGCC 1845
:::||||| :||| :|||
135 spAspAspGlyThrLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151
1846 GCAGAGACCGGACCGCTGCTTCCGCGGAGACAAATTTA.....AA 1889
||| ||||| :||| :|||
152 IleAlaAspSerThrLeuGlnGlyAlaGlyValArgValGluArgG 168
1890 CCGCAACATCAGCAAAACAGCGCAACTGTTTTCAGCGCG..... 1932
:::||||| :||| :||| :|||
168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHis 185
1933 .....AGACCAACACCGCGCTACATCATTTAAACGACAT 1971
||| :||| :|||
185 LeuGlyThrLeuGlnProLeu..... 191
1972 TGCTCGCAAAAGAGGCAATTCCTCGCGGGAATGCTGTGGACACGA 2021
||| ||| :||| :||| :|||
192 .....GlnProGlnsPneuProPserArgValValLeuGlyAspPh 206
2022 CTGATCAACCGC.....ACAT 2038
:::||||| :||| :|||
206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheVal 223
2039 TTAAGCGGAAACTTCATTAATTAAGCGGAGCGCGGTGTTCCCGC 2088
||| :||| :||| :||| :|||
223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisThrGlyGlyArg 239
2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGACAAATCAGCCCAAC 2138
:::||||| :||| :|||
240 AlaAlaGly.....ValAlaAl 245
2139 AGTTTGTGCTGCGACCGCATCAAGCCACACATCTGTACAGCTCG 2188
::: ||||| :||| :||| :|||
245 ameLAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGly 262
2189 AC..... 2190
||

```

262 spalaProAlaGlyGlyAlaValProGlyAlaValProGlyGlyAla 278
2191TGACGGGCTTGACAA 2207
279 ValProGlyGlyPheGlyProLeuLeuAspGlyTyrPtyGlyVal 293
2208 TTGTGTGAAAAAACCATTCAGCAGCATTAAGTATGCTTCATTTGACTA 2257
293 293
2258 AGACCGCAATCAGC...GCCAATGTGATTTGCCGATCAGCTCATTTA 2304
294AspValSerAspSerThrValAspLeuAlaGln 304
2305 AATCTCACAGGCTTGCCACACTCAAGCGCAATCTTAGTGCATAGGCCA 2354
305 SerIleValGlyAlaProGlnLeuGlyAlaAlaIleArgAlaGlyArgG 321
2355 TCACAGCTTATACAGTCAGCCCAAC.....GCCACCAAAAGCGCACG 2398
321 yAlaArgValThrValSerGlySerLeuSerAlaProHisGlyAsnV 338
2399 TTAGCCTCGGGGCAATGCCCAAGCAACATTTAAT 2433
338 aIleGlnThrGlyGlyGlyAlaArgArgPheProProAlaSerPro 354
2434 ...CAAGCCACATTTAAAGCGCAACACATCGCTTGCGCATGCT...TC 2477
355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaLeu 371
2478 ATTATCTTAAGCGACCGCGTACAAAAGCGCACTGACGCTTCCG 2527
371 uTyArgValLeuProGlnProValys.....LeuThrLeuAlaG 385
2528 GCACCGCTAAGGCAAGTAAAGCATTCGCACTCAAGGTAATGCTTCG 2577
385 GlyAlaGlnGly.....GlnIleAspIleVal 394
2578 CTAAGCGATAAGCGACTATTCATTTGAAGCAGCCGCTTACCGGACA 2627
395 AlaThr.....GluLeuProPheProGlyAla 404
2628 AATCAGCGCGCGCAAGATAGCGCATTTACCTTAAGACAGCGAATGA 2677
404 AserSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTyrT 420
2678 GCCTGCCGTCAGCGACGGAATTAGCAATTTAAACCTTGACAAAGCCAC 2727
420 hr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThr 435
2728 ATTACACTCAATTCGGCCTATGCGCACGATGCGGAGGCGCAAAACCG 2777
436 TrpValMet..... 438
2778 CAGTCGACAGATGCGCGCGCGCGCTTCGGCGCTTGCGGCGCTGCC 2827
439ThrAsp.....AsnSerAsnValGlyA 446
2828 TATTTATCCGTTACACCGCCAACTTGCGTAGAA..... 2859
446 IAlaLeuArgLeuAlaSerAspGlySerValAspPheGlnGlnProAlaGln 462
2860 ...TCCGTTTCAACAGCTGACGCTTAAGCGCAATTTGAACGCTGACGG 2906
463 AlaGlyArgPheLeuValLeuMetValAsp...ThrLeuAlaGlySerG 478
2907 AACATTCCGCTTATGTGGAACCTTTCGCTACCGCAGCAGCAAAATGA 2956
478 yLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspIleuV 495
2957 AGCTGGCGGAAGTTCCGAAGCGACTTACACTTGCGGCTGACAAATACC 3006
495 aIValMetArgAspAlaSerGlyGlnHisArgLeuTyrValArgAsnSer 511
3007 GGCACAGCAACTTCAAGACCTCGACACATTTAGCGTAGTGAGAAAGAA 3056
512 GlySerGlnProAlaSerGlyAsnThrMetLeuLeuValGln 525
3057 CAACAACCGCTGCTCGAAACCTTAATTTACCTGCAAAACGAA 3102
526 ...ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnIysAspG 541
3103 ...CAGCTGATGCGCGGCGGTGCTTACCAACTCATCGCAAGACGCG 3150
541 IlyValAlaSplIleGlyThrTyArgTyArgLeu 552
3151 GAATTCCGCTTCATTAATCCGTCAAAGACAGAGCTTTCGCAAACT 3200
552 552
3201 CGGCAAGCGCAGAGCAAAACAGCGGAAACAGACAGCGCAAGCC 3250
553AlaAlaAsnGlyAsnGlyGlnTyr. 560
3251 TTAGCGCTGATTCGCGCGCGCGCGATGCGTGGAAAGACAGAAAGC 3300
561 ...SerLeuValGlyAla..LysAlaProProAla.....ProLysPr 573
3301 GTTCCCGAAGCGCGCGCGCGCAGGCGGGAATGTCGCATTTATGCA 3350
573 OAlaPro...GlnProGlyProGlnProGlyProGln 584
3351 GCGGAGGAGAAAGAAACAGGTCGACGCGGATTAAGACACCGCTTCG 3400
585ProPro... 586
3401 CGAAACAGCGGAAGCGGAACCGCGCGCTACACCGCTTCCCGC 3450
587GlnProProGln..ProProG 593
3451 GCCCGCGCGCGCGCGGATTTCCCGCACTGCAACCCCA 3492
593 IlnProProGlnProPro..GlnProProGlnArgIlnProGlnAlaProAl 609
3493 ...CGCAGCGCCCAACCGCAGCGCAGCATGATCAGCGCTTATGCGCAATGCG 3541
609 aProGlnProProAlaGlyAlaArgGlyLeu...SerAlaAlaAlaAsnAla 625
3542 GTTGAAGTGAATTTCCGCGCAGCTCAACAGCTT.....TTCGCGCTA 3585
625 IAlaAlaAsnThrGlyValGlyValGlyLeuAlaSerThrLeuTyrGlyAlaGln 641
3586 CAGACGCAATTTAGACCGCTATTTGCCGAGACCGCGCCGCA 3627
642 SerAsnAlaLeuSerIlyAspGlyGlyIleLeuArgLeuAsnProAla 658
3628 ...GCCGTTTGGACAAAGCGCATTCGCGCAGCAACCACTACCTGCTGCG 3673
658 aGlyGlyAlaTyrGlyArgGlyPheAlaGlnGlnGlnLeuAspAsnAla 675
3674 AAGATTCCGCGCGCTACCGGCCAACAACGCACTGCGGCCAATCGTAG 3723
675 rGlyAlaGlyArgArgPheAspGlnIlyValAlaGlyPheGlyLeuAla 691
3724 CAGAAAAACCTC.....GGCAGCGCGCGCGTGGCGCATCGCTG 3761
692 AspHisAlaValAlaValAlaGlyAlaGlyArgTyrHisLeuGlyLeuAl 708
3762 TTGCGCAACCGCGCAAGCAACCTTC...GACGAGCGCATGCGCACT 3808
708 aGlyTyTyrArgGlyAspArgGlyPheThrGlyAlaGlyGlyGlnHisT 725
3809 CGGCAAGCGCTTGGCGCAGCGCGCTTTCGGGCAATGAGCGCATGCAAG 3858
725 hnsPheValHisValGlyGlyTyArgAlaThrTyTyrIleAlaAsnSerGly 741

291 eTyrleu.....GlyLeuSerGluValThrPheAspIleAlaasp 304
1450 GGTACAGTATTTGGATCAGCAGGACGATTAAGCAAAACAAAC 1499
||| : : : : : ||| : : : : :
305 GlyLeuThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspse 321
1500 CTTTAACTGAATCGCTGGTC.....AGCGCAGGGGACGGTCACAC 1543
: : : : : ||| : : : : :
321 rIleAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValI 338
1544 TGAATCCCATAT..... 1557
||||| : : : : :
338 euAsnAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGly 354
1557 1557
355 GluValThrLeuGlyLysGlySerAsnSerLeuMetAsnValGlyAspThrH 371
1558CAGTTCAACCCCGCAAACTATTTCCGCTTCCGGCGGACGTT 1603
||| : : : : : ||| : : : : :
371 scysGlnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerI 387
1604 TGGAT.....TTAAC..... 1614
: : : : : ||| : : : : :
387 LeuAspLysTyrGlnAsnGlnAlaGlyLeuAsnValGlySerThrGlnGln 403
1615GGCATTCGCTTCTTCACACGCT.....ATTCAAAATACCGA 1652
: : : : : ||| : : : : :
404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420
1653 TGAAGGGGCGATGATTCACACCAATTCACAAAGCAAAAGAACCCGTTA 1702
| : : : : : : : : : : :
420 PAlaGlyGlyAsnValThrValAsnGlnGlySerPheAlaGlyThrIleG 437
1703 CCATTACAGGCAATTAAGATATCTTACACCGCGCATTAACACACGCTG 1752
: : : : : ||| : : : : :
437 IuGlyAlaGlyLysLeuThrIleAlaGlnAsnGly...SerTyrValIle 452
1753 GATTACCAAAAAAGAAATTCCTTACACGCGTGGTTGGCGAAAGATAC 1802
: : : : : ||| : : : : :
453 AlaglyAlaGlnSerMetAlaLeuThrGly.....AspIle 464
1803 GACCAAAAGCAACGGCGGTCAACCTGTTTACACGCCCGCCGA.... 1848
: : : : : ||| : : : : :
464 euAlaValAspAlaGlyAlaValLeuSerLeuGlnGlyAspAlaAlaAsp 481
1849GAAGACGCGACCTGCTGCTTCCGCGGA... 1878
481 euAlaAlaLeuGlnAspAspProGlnSerIleValLeuAsnGlyGlyMet 497
1879 ACAATTTAAACGGCAGCATCAGCAACAAACGCAAACTGTTTTCAG 1928
: : : : : ||| : : : : :
498 LeuAspLeuSerAspPheSerThrTyrPoleSerGlyThrSerTyrLysAs 514
1929 CGGACAGCAACACCGCACCGCTAC..... 1953
: : : : : ||| : : : : :
514 PglyLeuGlnValSerGlySerGlyThrValIleGlySerGlnAspV 531
1954AATCATTTAAACGACCATTTGGTCGCAAAA 1983
531 AlValAspLeuAlaGlyGlyAsnAspMetHisIleGlyGlyAspGlyLys 547
1984 GAGGGC.....ATTCCGCGGGGGA.....ATCGT 2009
: : : : : ||| : : : : :
548 AspGlyValTyrValValIleAspAlaGlyAspGlyGlnValSerLeuAl 564
2010 GTGGGACACGACTGATCAACCGCACATTTAAAGCGAAACCTCCAAA 2059
: : : : : ||| : : : : :
564 AsnAspAsnGlnTyrLeuGlyThrThr.....GlnI 575
2060 TTAAGGCGGACAGCGCGTGTTCGCGCAATGTTCGCAAAAGTGAAGGC 2109
||| : : : : : ||| : : : : :
575 IeAlaSerGlyThrLeuMetValSerAspAsn.....SerGlnLeuGly 589

2110 GATTGCAATTTGACCAATCAGCCCAACAGTATTTTGGTGTGCGACCGCA 2159
||| : : : : : ||| : : : : :
590 TyrThrHisTyrAsnArg.....GlnValIlePheThrAspLysProG 604
2160 TCAAGC.....CACCAATCTGTACAGTTCGGACGTGA 2194
: : : : : ||| : : : : :
604 nGluSerValMetGluIleThrAlaAsnValAspThrLysThrThr 621
2195 CGGGTCGACAAATGTGTGCAAAAAACCATTTACCGACGATTAAGTAT 2244
||| : : : : : ||| : : : : :
621 hrGlnHisGlyArgAspIleGlu.....MetAlaGlyAspGlyVal 635
2245 GCTTCATTTAGTAAGACGACATCGACGGGCAATGTGATCTGGCGATCA 2294
||| : : : : : ||| : : : : :
636 AlValAspAlaGlyValAspThrGlnTyrPoleLysLeuMetAlaAspse 652
2295 C.....GCTCATTTAAATCTCAAGGCGCTTG 2320
652 rSerGlyGlnHisGlnAspGlnGlySerThrLeuThrLysThrGlyAlaG 669
2321 CCACACTCAACGGCAATCTTAGTCAATGAGCGATACGTTATACGTC 2370
: : : : : ||| : : : : :
669 LysThrLeu.....GluLeuThrLysSerGlyThrThrGlnSerAlaVal 683
2371 AGCCACAACGCCACCCAAACGGCAACCTTACGCTGCGGCAATGCCA 2420
: : : : : ||| : : : : :
684ArgValGlnGlnGlyThrLeuGln.....GlyAspValAl 695
2421 AGCAACATTTAATCAAGCCACATTA.....AACGCAACACAT 2458
||| : : : : : ||| : : : : :
695 AspIlePheThrTyrAlaSerSerLeuThrValGlyAspGlyAlaThrP 712
2459 CGGCTTGGGCAATGCTTCAATTAATCTAAGCAGCAGCGGTACAAAAC 2508
: : : : : ||| : : : : :
712 heValIThrGly.....AlaAspGlnAspIleGlnSer 722
2509 GCGAGTCGACGCTTCCGGCAACGCTAAGCAACGTAAGCATTCGCCG 2558
: : : : : ||| : : : : :
723 IleAspAlaThrSerSer..... 728
2559 ACTCAACGTAATGTCTCCCTACCGATAAGCAGATTCGATTTGA 2608
||| : : : : : ||| : : : : :
729GlyThrIleAspIleSerAspGlyThrValLeu..... 739
2609 GCAAGCGCTTACCGGCAAAATCAGCGCGGCAAGATACGGCATTAAC 2658
||| : : : : : ||| : : : : :
740ArgLeuThrGlyGln.....AspThrSerValAla 749
2659 TTAAGACACG.....GAATGAGCGCTGCCGTCAGCGACGCA 2696
||| : : : : : ||| : : : : :
750 LeuAsnAlaSerLeuPheAsnGlyAspGlyThrLeuValAlaAlaThrAs 766
2697 A.....TTAGCAATTTAAAC..... 2712
||| : : : : : ||| : : : : :
766 PglyValThrLeuThrGlyGlnLeuAsnThrAsnLeuGluThrAspSerL 783
2713CTGACACGCGCACCATTTACACTGATTCGCTATGCGCAC 2754
||| : : : : : ||| : : : : :
783 euThrTyrLeuSerAsnValThrValAsnGlyAsnLeuThrAsnThrSer 799
2755 GATGCGGAGGGGCGCAACCGGACGTGCGACAGATCGCGCGCGCG 2804
||| : : : : : ||| : : : : :
800 GlyAlaValSerLeuGlnAsnGlyAlaGlyAsp..... 811
2805 TTCGCGCGCTTGGCGCGTTCCTTATTCGTTACCGCGCAACTTGG 2854
811 811
2855 TAGAATCCGTTTCAACACGCTGACGTTAAAGCGCAATTTGACGCTGAG 2904
||| : : : : : ||| : : : : :
812ThrLeuThrValAsnGlyAspTyrThrGly 822

A:Molecule type: DNA
 A:Residues: 1-1250 <STO>
 A:Cross-references: GB:AE005174; NID:g12516568; PIDN:AA57362.1; GSPDB:GN00145; UNGP:234
 A:Experimental source: strain 0157:H7, substrain EDL933
 A:Genetics:
 A:gene: yfai

alignment_scores:

Quality: 295.00 Length: 1446
 Ratio: 0.452 Gaps: 73
 Percent Similarity: 45.159 Percent Identity: 20.678

alignment_block:

US-09-303-518d-649 x F85862 ..

Align seg 1/1 to: F85862 from: 1 to: 1250

```

637 ACCGGCAAGTTCATATCATATTCAGTGGCTATCT..... 675
46 SerArgInSerLeuSerGlyIleThrGlnValTrpSerIleAlaSpI 62
676 ....TGCTCGTGTGGCAATACCTTTGCACAAATGATCAGGCTG 721
62 yGIInTrpLeuValPheSerAsp...MetThrAsnAlaSerGlyAla 78
722 GCACACTCAACTTAGTAGTGAATAAATTAAACATAGCCCATGCTTT 771
78 Laval.....Phe 80
772 TTACCAAGAGAGGCTCATTT.....GGCAGACAGTGGCTC 806
81 LeuGlnGlnGlyAlaGluPheThrLeuSerProGluAsnGluThrClyme 97
807 ACCAATGTTTATCTATGATGCCCCAAAAGCAAAAGTGTAAATTAATGGGG 856
97 tThrLeuPheAlaAsnAsnThr.....ValSerGlyG 108
857 TATGCAAAAGGGCAACCCCTTATATAGAAAAGCAATGGCTTCAGCTG 906
108 lutyAsnAsnGlyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124
907 GTTCGTAAGATTGGTTATGATGAATCTTGTGCGATACCATTC 956
125 LeuThr.....AspValIlePheSerGlyAsnValAlaGlu 136
957 ACTATTCTAGCAACCACTCAAAATGGCAATATCTTTTACGACGATA 1006
136 yGlyTyr.....GlyGlyAlaIleTyrSerSerGlyT 147
1007 ATAAATGGCACAGAAAATCAATGCCAATGAAACAC..... 1044
147 ThrAsnAspThrGlyAlaIleAsnLeuAsnValThrAsnAlaValPheArg 163
1045 AATTCCTGCGCTAATAGATTAAACACGACCGCTTCAATGTTAAT... 1092
164 AsnAsnIleAlaAsnAspGlyLysGlyAlaIleTyrThrIleAsnAs 180
1093 ....GTTCTTTATCCGAGACAGCA.....AGAGAACCTGTTTATCATC 1132
180 nasPileTyrLeuSerAspAspValPheAsnAsnAsnGlnIleTyrThrS 197
1133 CTGCAGTGTGTGTCACAGTTATGACCCACCACTCAATATGAGAAAT 1182
197 erThrSerTyrSerAspGlyAspGlyAlaIleAspValThrAspAsn 213
1183 ATTTCCTTTATTGACGAGAAAGCGCAATTCATCT..... 1221
214 AsnSerAspSerLysHisProSerGlyTyrThrIleIleAsnAsnThrAl 230
1222 ....ACCGCAACATCAATCAAGTGGTGGATTTATTTTCCAGAGAG 1267
230 aPheThrAsnAsnThrAlaGlnGlyTyrGlyAlaIleTyrThrAsnS 247

```

```

1268 ATTTTACG.....GTCTCCCTGAAAATTAACCAACT 1299
247 eValAlaThrAlaProTyrLeuIleAspIleSerValAspAspSerTyrSer 263
1300 TGGCAAGGGCGGGCTTCATATCAGTGAAGACACTACCGTTACTTGCA 1349
264 ...GlnAsnGlyAlaValLeuValAspGluAsnAsnSerAlaIleGlyTyr 279
1350 AGTAAACGGCGTGGCAAAAGCAGCCGCTGTCAAATTCGCAAAAGCGCAGC 1399
279 rGlyAspGlyPro.....SerSerIleAlaGlyGlyPheM 291
1400 TGCAGCTTCAAGCCAAAGGGCAAAACAGGCTGCATCAGGCTGGCGAC 1449
291 eTyrLeu.....GlyLeuSerGluValThrPheAspIleAlaAsp 304
1450 GGTACAGTCATTTGTGATCAGCAGCAGACGATTAAGCAAAACAAACAGC 1499
305 GlyLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSe 321
1500 CTTTAGTGAATTCGGCTGGTC.....AGCGCAGGGGTACGGTGCAC 1543
321 rIleAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValL 338
1544 TGAATGCCGATAT..... 1557
338 euAsnAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGly 354
1557 ..... 1557
355 GluValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHI 371
1558 ....CAGTTCAACCCGCAACAACTCTATTTGCGTTTCGGCGGACGTT 1603
371 scysGlnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerI 387
1604 TGGAT.....TTAATC..... 1614
387 leAspLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403
1615 ....GGCATTCGCTTTCGTTCCACCGT.....ATTCAAAATFACGA 1652
404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420
1653 TGAAGGGCGCGATGTGTCAACCAATCAAGACAAAGAAATCCACCGTTA 1702
420 palaglyGlyAsnValThrValAsnGlnGlySerPheAlaGlyThrIleG 437
1703 CCATTACAGCATTAAGATATTGCTTACACCGGCAATACACAGCTTG 1752
437 luyGlyAlaGlyGlnLeuThrIleAlaGlnAsnGly...SerTyrValLeu 452
1753 GATGCAAAAAAGAAATGCTTACAAACGTTGGTTTGGCGAGAAGATAC 1802
453 AlaglyLysIleGlnSerMetAlaLeuThrGly.....AspII 464
1803 GACCAAAAGCAAGGGGGGCTCAACCTGTTTACACCCGGCGCA.... 1848
464 eValValAspAlaGlyAlaValLeuSerLeuGlnGlyAspAlaAlaAspL 481
1849 .....GAAGACCGACCCCTGCTGCTTCGGCGGCA.... 1878
481 euAlaIleAlaLeuGlnAspAspProGlnSerIleValLeuAsnGlyGlyMet 497
1879 ACAAAATTTAAAGCGCAACATCACGCAAAACAGCGCAAACTGTTTTTCAG 1928
498 LeuAspLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLysAs 514
1929 CGGCAACCAACACCGCAGCCCTAC..... 1953
514 pGlyLeuGluValSerGlySerSerGlyThrValIleGlySerGlnAspV 531

```

```

1954 .....ATCATTTAAACGACCATTTGGTCGCAAAA 1983
531 aValaspLeuValaglyGlyasnAspMetHisIleglyIysaspGlyLys 547
1984 GAGGCG.....ATTCCTCCGGGGGA.....ATCGT 2009
548 AspGlyValIytrValIValIleaspAlaGlyaspGlyIValSerIeuVal 564
2010 GTGGGCAACAGACTGGATCAACCGCACATTTAAAGCGGAAACCTTCCAAA 2059
564 aasnAspasnGlnIytrLeuGlyThr.....GlnI 575
2060 TTTAAAGCGGACAGCGCGTGTTCGCCCAATGTGCCAAAGGAAAGGC 2109
575 leaIleSerGlyThrLeuMetValSerAspAsn.....SerGlnLeuGly 589
2110 GATGGATTTGGCAATCAGCCAGCGAGTTTGGTGTGCGACGCA 2159
590 TyrThrIstIytrAsnArg.....GlnValIlePheThrAspIySproG 604
2160 TCAAAAGC.....CACACATCTGTACAGCTTGGACTGGA 2194
604 ngluSerValMetGluIleThrAlaAsnValAspThrArgSerThrThrT 621
2195 CGGGTGTGACAAATGTGTGGAATAAACCAATCCGACGATTAAGTGAT 2244
621 hrGlnHisGlyArgAspIleGlu.....MetArgAlaAspGlyGluVal 635
2245 GCTTCATTTAGTAAAGCAGCATCAGCGGAGCATGTGATCTTGGCGATCA 2294
636 AlaValaspAlaGlyValAspThrGlnIytrPheGlyAlaLeuMetAlaAsp 652
2295 C.....GCTCATTTAAATCTCACAGGGCGTTG 2320
652 rSerGlyGlnHisGlnAspGlySerThrLeuThrIySthrGlyAlaG 669
2321 CCACACTCAAGCGCAATCTTAGTCAAAATGCGATACACGTTATACATC 2370
669 IyThrIleu.....GluLeuThrIleSerGlyThrGlnIserAlaVal 683
2371 AGCCCAACAGCCACCCAAACGCAACCTTAGCCTCGTGGCAATGCCCA 2420
684 .....ArgValGlnGluGlyThrLeuGln.....GlyAspValAla 695
2421 AGCAACATTTATCAAGCCACATTA.....AAGCGCAACACAT 2458
695 aAspIlePheProIytrAlaSerSerLeuIytrValGlyaspGlyAlaThrP 712
2459 CGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACGACCGCGTACAAAAC 2508
712 heValThrGly.....AlaAspGlnAspIleGlnSer 722
2509 GGCAGCTGAGCGCTTCCGGCAGCGCTAAGCAACGTAAAGCATTCGCG 2558
723 IleAspAlaThrSerSer..... 728
2559 ACTCAACGGTATGTCTCCCTAGCCGATAAAGCAGTATTCATTTTGAAA 2608
729 .....GlyThrIleAspIleSerAspGlyThrValIeu..... 739
2609 GCAGCGCGCTTTCAGGACAAATCAAGCGGCGGACGATAGCGCATTAAC 2658
740 .....ArgLeuThrGlyGln.....AspThrSerValAla 749
2659 TTTAAAGACAGC.....GAATGACGCTGCGCGTACGAGCGGA 2696
750 leuAsnAlaSerIeuPheAsnGlySaspGlyThrIeuValAsnAlaThrAs 766
2697 A.....TTAGCAATTTAAAC..... 2712
766 pGlyValIyThrIeuThrGlyGlyIeuAsnThrIeuGluThrAspSerL 783
2713 .....CTTGACAAAGCCACCATTTACATCAATTCGCGCTATCGCCAC 2754

```

```

783 eutIyThrIeuSerAsnValIytrValAsnGlyAsnLeuThrAsnThrSer 799
2755 GATGGCGGACGGCGCAAAACCGGACGTCCGACACATCGCGCGCGCGCG 2804
800 GlyAlaValSerIeuGlnAsnGlyValAlaGlyAsp..... 811
2805 TTCCGCGCGCTTCCGCGCGTCCCTATTATTCGTTACACCGCCAACTTCG 2854
811 ..... 811
2855 TAGAATCCCGTTTCAACACGCTGACGTAACGCAATTTGAACGCTCAG 2904
812 .....ThrIeuThrValAsnGlyAspIyThrGlyGly 822
2905 GGAACATTCGCGCTTATGTCCGAATCTTCGCGC.....TACCGACGGA 2948
823 GlyThrIeuLeuAspSerIeuIeuAsnGlyAspAspSerValSerAs 839
2949 CAATTTGAAGCTGGCGGAAAGTTCCGAAGGACATTACACTTGGCGGTCA 2998
839 pGlnIeuValMetAsnGlyAsnThrIleGly***ThrThrValValAla 856
2999 ACAAT...ACGCGC...AACGAACTGCAAGCCTCGACAAATTTGACGCTA 3042
856 snSerIleThrGlyIleGlyGluProIySerThrGlyIleIyValAla 872
3043 GTGGAAAGAAAGACAAACAAACCGCTGTCGAAACCTTAAATTTCCCT 3092
873 AspPheIleAlaIAspProIyThrIleGlnAsnAlaGlnPheSerLe 889
3093 GCAAAACGAA...CACGTGATGCGCGCGCGCTGACCACTATCAC 3139
889 uIaGlySerGlyTyValAsnMetGlyAlaIytrAspIyThrIeuValG 906
3140 GCAAAACGCGGAGTTCGCCCTGCATATCCGCTCAAGAACAAGAGCTT 3189
906 IuAspAsnAsnAspIyTyLeu..... 913
3190 TCCGAAAACTCGCAGACAGCAAGCAAAAAACAGCGCGGAAAAAGACAA 3239
913 ..... 913
3240 CGCGCAAAAGCTTTAGCGCGCTGATTTGGCGCGGCGCGATGCGTCGAAA 3289
914 .....A 914
3290 AGACAGAAAGCGTTGCCGAACCGCGCGCGAGCAGCGCGGGAATAATGTC 3339
914 rSerGlnGluValIyThrProPro..... 921
3340 GGCATTATGAGGCGGAGGAGAAAGAAAAACGGGTCCAGCGGATTAACA 3389
921 ..... 921
3390 CACCGCGCTTGGCAAAACAGCGGAGGAAACCGCGCGGTACACCG 3439
922 .....SerProIyAspProIyThrP 930
3440 CCTTCCCGCGCGCGCGCGCGCGCGCGGAGTTTGGCGCAATGCAACCC 3489
930 roAspProIyAspProIyThrGlnAspProIyAspProIyAspProIyPro 946
3490 CAACCGGACCCCAACCGCAGCGCGACCTGATCAACCGCTTATGCCAATAG 3539
947 ThrProIaIytrGlnProValIeuAsn..... 955
3540 CGGTTTGAAGTAATTTCCGCCCAAGCTTCACAGCGTTTGGCGGTACAG 3589
956 .....AlaIySValGlyIyTyIleuAsnIleuAlaIAla... 968
3590 ACGAATTAGACCGCGCTATTGCGGAAAGACCGCGCAACGCGGTTTGACA 3639

```

```

969 .....AsnGlnAlaPheMetMetGluArgArgAspHisAlaGlyGly 982
3640 AGCGGC.....ATCGGGACACCAACACATCCG 3668
983 AspGlyGlnThrLeuAsnLeuArgValIleGlyGlyAspTyrHisTyrTh 999
3669 TTGGCAAGATTTCGGCGCTACGCCCAACAAACGACCTGCGCAATTCG 3718
999 rAlaAlaGlyGlnLeuAla.....GlnHisGluAspThrSerThrValG 1014
3719 GTATCGCAAAACCTCGGACGGGGCGC..... 3747
1014 LntLeuSerGlyAspLeuPheSerGlyArgTyrGlyThrAspGlyGlyTyr 1030
3748 ...GTGCGCATCCCTG.....TTTTCGCAACCGGACCGCAAAACACCTT 3788
1031 MetLeuGlyIleValGlyGlyTyrSerArgAsnGlnGlyAspSerArgSe 1047
3789 CGACGACGCGATCGGCACTCGGACGCGCTTGGCCACGCGCGCTTTTCG 3838
1047 rSerMetThrGlyThrArgAlaAspAsnGlnAsnHisGlyTyrAlaValG 1064
3839 GG.....CAATACGATCGACAGGTTTCAATCGCG 3870
1064 LyleuThrSerSerTyrPheGlnHisGlyGlyGlnLys..... 1076
3871 ATCAGCGCGGCGCG.....GGTTTTCAGCAG 3896
1077 .....GlnGlyAlaTyrLeuAspAsnTyrPleuGlnTyrAlaTyrPheSe 1091
3897 CGGACGCTTTCAGACGCGCATCGAGGCAAAATCCGCGCGCGCTGCTTCG 3946
1091 rAsnAspValSerGlnHisGluAspGly.....ValAspH 1103
3947 ATTAC.....GGCATTCGACGATACCGCGCGCTTTC..... 3981
1103 IeTyrHisSerSerGlyIleIleAlaSerLeuGlnAlaGlyTyrIleTyr 1119
3982 .....GGCGATTTCGCGATCGACCGGACATCGCGCGCAACCGCGTA 4022
1120 LeuProGlyArgGlyValIleGlnProGlnAlaGlnValIleTyrGln 1136
4023 TTTCGTCGCAAAAGCGGATACCGCTACGAAAC..... 4056
1136 nGlyValGlnGlnAspAspPheThrAlaAlaAsnArgAlaArgValSerG 1153
4057 .....GTCAATATCGCCACCGCGCGCTTCATTCACCGCTAC 4095
1153 LntSerGlnGlyAspAspIleGlnThr..... 1161
4096 CGCGCGCGCATTAAGCGACGATTAATTCATCAACCGCGCAACACATTCG 4145
1162 ArgLeuGlyLeuHisSerGlnTyr.....ArgThrAlaValHis..... 1174
4146 CATCACGCGCTATTGAGCTGTGC...TATACGATCGCGCTTCGGGCA 4192
1175 ...ValIleProThrLeuAspLeuAsnTyrTyrHisAspProHisSerThrG 1191
4193 AAGTCGCAACGCGCTCAATACCGCGCTATTCGCTCAGGATTCGGGCA 4242
1191 LntIleGlnGluAspAlaSerThr.....IleSerAspAspAlaValLys 1205
4243 ACCGCG...AGTCGGAATGGGGGTAAACCGCAATCAAGGTTTCAC 4289
1206 GlnArgGlyGlnIleLysValGlyValThrGlyAsnIleSerGln...Ar 1221
4290 GCTGTCCCTCCACGCTGCGCGCGC.....AAAGCGCGCAACTGGAG 4333
1221 gValSerLeuArgGlySerValAlaIleTyrGlnLysGlySerAspSph 1238
4334 CGCAGACACGCGCGCATCAATTAAGCTACGCTGG 4371
1238 LntGlnThrAlaGlyPheLeuSerMetThrValLysTyr 1250

```

```

seq_name: p1r2:A32560
seq_documentation_block:
outer membrane protein P.69 precursor - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: A32560
R:Charles, I.G.; Dougan, G.; Pickard, D.; Chatfield, S.; Smith, M.; Novotny, P.; Mori
Proc. Natl. Acad. Sci. U.S.A. 86, 3554-3558, 1989
A:Title: Molecular cloning and characterization of protective outer membrane protein
A:Reference number: A32560; MIMD:89264462
A:Accession: A32560
A:Molecule type: DNA
A:Residues: 1-910 <CHAS>
A:Cross-references: GB:J04560; NID:g144053; PIDN:AAA22980.1; PID:g144054
A>Note: It is uncertain whether Met-1 or Met-3 is the initiator
C:Keywords: membrane protein
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-910/Product: outer membrane protein P6.9 #status predicted <MAT>

alignment_scores:
Quality: 283.00 Length: 1100
Ratio: 0.581 Gaps: 52
Percent Similarity: 44.273 Percent Identity: 21.909

alignment_block:
US-09-303-518D-649 x A32560 ..
Align seg 1/1 to: A32560 .from: 1 to: 910

1366 AACGACCGCGCTGTCCAAATCGGCAAGGACCGCTGCAGCTTCAGGCCAA 1415
||||| : : : ||| ||| : : : : : : : : : : : : : : : : : : : :
38 AsnGlnSerIleValIleTyrGlnArgGlnHisGlyLeuHisIleGlnIleG 54
1416 AGGGGAAACCAAGGCTCGATCAGCGCTGGCGGACGAGTACAGTATTGG 1465
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
54 nGlySerAspProGlyValArgThrAlaSerGlyThrThrIle.... 69
1466 ATCAGACGCGACGATTAAGGCAAAACAGCCTTACTGAATTCGGC 1515
69 ..... 69

1516 TTGGTCAGCGCGGAGGTACCGTGCAATCGAATGCCATATGAGTTCAA 1565
||||| : : : : : : : : : : : : : : : : : : : : : : : :
70 LysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGlnLys..... 83
1566 CCCGACCAACTCTATTTCGCTTTCGCGGCGGAGCGTTTGATTTAAGC 1615
84 ....ProAlaAlaGlnLeuGlnPheArgAsnGlySerValThrSerSerG 99
1616 GCGATTCGCTTTCGCTTCACCGGATTCAAATACGATGAAAGGCGGATG 1665
||| : : : : : : : : : : : : : : : : : : : : : : : :
99 Lys.....GlnLeuSerAspGlyLeu... 106
1666 ATTGTCAACCAATCAAGCAAGAATCCACCGCTTACCATTTACAGGCAA 1715
107 .....ArgArgPheLeuGlyThrValThrValLysAlaGln 118
1716 TAAA.....GATATTGCTTCAACCGGCAATTAACACAGCTTGATTA 1756
||||| : : : ||| ||| : : : : : : : : : : : : : : : : :
118 LysLeuValAlaAspHisAlaThrLeuAlaAsnValGly..... 131
1757 GCAGAAAAGAAATTCCTACACGCGTTGGTGGCGAGAAAGATACGACC 1806
132 .....AspThrTyr 134
1807 AAACGACGCGCGGCTCAACCTGTT.....TACGACCGCGC 1844
||| : : : ||| : : : : : : : : : : : : : : : : :
135 AspAspAspGlyIleAlaLeuTyrValAlaAlaGlyGlnAlaGlnAlaSe 151
1845 CGCAGACGCGACCGCTGCTTCGCGCGGAAACAAATTTA.....A 1888

```



```
404 ThrPheValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu..... 417
1674 CCACATCAAGACAAAGATCCACCCGTACCAT..... 1707
418 ...AsnIleAspAlaGlyIleGlnSerMetAlaLeuThrGly..... 433
1708 .....ACAGCAATTAAGATATGCTACACCGCAAT 1740
433 IagIleIleIleGlnGlyGlnLeuThrIleAlaGlnAsnGly... 448
1741 AACACAGCTTGATAGTACAAAGAAATTCCTACACGGTTGGTGG 1790
449 SerTyrValLeuAlaGlyAlaGlnSerMetAlaLeuThrGly..... 462
1791 CGAAGAAATACGACCAAAACGAGCGCTCAACCTTTGATACAG 1840
463 .....AspIleValValAspAspGlyAlaValLeuSerLeuGlnGlyA 477
1841 CCGCCGCA.....GAGACCGCACCCCTGCTGCTT 1869
477 spAlaIlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeu 493
1870 TCCGCGCGAACA...AATTAAAGCGCACATCCGCAACAACGCGCA 1916
494 AsnGlyGlyValLeuAspLeuSerAspPheSerThrTrpGlnSerGly 510
1917 ACTGTTTTCAGCGCGACACACACCGCACCGCTAC..... 1953
510 rSerTyrAsnAspGlyLeuGlnValSerIleSerSerGlyThrValIleG 527
1954 .....AATCATTTAAACGCCAT 1971
527 IySerGlnAspValValAspLeuAlaGlyIyAspAsnLeuHisIleGly 543
1972 TGGTCGCAAAAGAGGCGCAT.....CTCG 1997
544 GlyAspGlyLysAspGlyValIyValValAlaAspAlaSerAspGly 560
1998 CGGGGAATCGTGGGACACAGCATGATCAACCGCATTTAAAGCG 2047
560 nValSerLeuAlaAsnAsnSerTyrLeuGlyThr..... 573
2048 AAACCTCCAATTAAGCGGACAGCGGCTTCCGCAATGTGGC 2097
574 .....GlnIleAlaSerGlyThrLeuMetValSerAspAsn..... 585
2098 AAGTGAAGCGCATTTGCGATTTGAGCAATCAGCCCAAGCGTTTGG 2147
586 SerGlnLeuGlyAspThrHisTyrAsnArg.....GlnValIlePheThr 600
2148 TGTCCGACCGCATCAAGC.....CACACAATCTGTACAC 2182
600 rAspIyGlnGlnGlnSerValMetGluIleThrSerAspValAspThrA 617
2183 GTTGGAGCGAGCGGCTGTGACAAATTTGTGCAAAAAACCATTCAGC 2232
617 rGSerAspAlaAlaGlyHisGlyAspIleGlu.....MetArgAla 631
2233 GATAAAGTATGATCTTCATGATGATAGACGACATCAGCGCATGTGCGA 2282
632 AspGlyGluValAlaValAlaAspAlaGlyAlaSprHisGlnTrpGlyAlaLe 648
2283 TCTTCCCATCAC.....GCTCATTTAAATC 2308
648 uMetIlaAspSerSerGlyGlnHisGlnAspGluGlySerThrLeuThrL 665
2309 TCACAGGCTTGCCGACACTCAACGGCAATCTTAGTGAATGGGATACA 2358
665 ystHisGlyAlaGlyThrLeu.....GluLeuThrAlaSerGlyThr 679
2359 CGTATATACGTACGACCAACGCCAACCCAAACGCAACTTAGCTGTGT 2408
::: :::::
680 GlnSerAlaVal..... 683
2409 GGGCAATGCCCAAGCACATTTAATCAAGCCCATTAACGGCAACACAT 2458
684 .....ArgValGlnGlnGlyThrLeuGlySerGlyAspValA 695
2459 CGGCTTCGGGCAATCTTCATTTAATCTAAGCGCACCGCC..... 2499
695 Ia.....AspIleLeuProTyrAlaSerSerLeu 704
2500 ...GTACAAACGGCAGTCTGACGCTTCGCGCAACGCTAAGCGCAACGT 2546
705 TrpValGlyAspGlyAlaThrPheValThrGlyAlaAspGlnAspIleG 721
2547 AAGCATTCGCGACATC...AACGTAATGTCTCCCTACCGATTAAGCGAG 2593
721 nSerIleAspAlaIleSerSerGlyThrIleAspIleSerAspGlyThrV 738
2594 TATTCATTTGAAGCAGCGCTTACCGGACCAATACGGCGGCGCAG 2643
738 AlLeu.....ArgLeuThrGlyGln..... 744
2644 GATACGGCATTAACCTTAAGACAGC.....GAATGACGCT 2681
745 AspThrSerValAlaLeuAsnAlaSerLeuPheAsnGlyAspGlyThrLe 761
2682 GCCGTCAGGACGGA.....TTAGCAATTTAAAC..... 2712
761 uValAsnAlaThrAspGlyValThrLeuThrGlyGlnLeuAsnThrAsnL 778
2713 .....CTTGACAACGGCACCATTACACTCAAT 2739
778 euGlnThrAspSerLeuThrTyrIleuSerAsnValThrValAsnGlyAsn 794
2740 TCCGCTATTCGCCAGATCGCGAGCGGCGCAACCGCAAGTGGACAGA 2789
795 LeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValAlaGlyAs 811
2790 TGGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTTATTCGTTA 2839
811 P..... 811
2840 CACGCGCAACTTCGTTGAATCCGTTTCAACAGCGTGACGTAACGCG 2889
812 .....ThrLeuThrValAsnGly 817
2890 AAATTAACGCTCAGGACATTCGCTTATGTGCAACTCTTCGCG... 2937
818 AspTyrThrGlyGlyGlyThrLeuLeuLeuAspSerGlyLeuAsnGlyAs 834
2938 ...TACCGCAGCAGCAAAATTTGAAGCTGCGGAAAGTTCCGAAGGCACT 2983
834 pAspSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsnT 851
2984 ACACCTTGGGCGTCAACAT...ACGCGC...AAGCAACTGCAAGCGCTC 3027
851 hrThrValValAlaAsnSerIleThrGlyIleGlyGlnProThrSerThr 867
3028 GAACAATTCAGCTAGTGAAGAGAAAGACACAAACCGCTGTCCGAAAA 3077
868 GlyIleLysValAlaAspPheAlaIlaAspProThrGlnPheGlnAsnAs 884
3078 CCTTAATTCACCTGCAAAAGAA...CAGTCGATTCGCGGCGCTGCG 3124
884 nAlaGlnPheSerLeuAlaGlySerGlyTyrValAlaMetGlyAlaTyrA 901
3125 GTTACCAACTCATCCGCAAGACGCGAGTTCGCTCGATATACGCGGTC 3174
901 sPtyrThrLeuValGluAspAsnAsnAspTyrIleu..... 913
3175 AAGAACAAGAGCTTTCGACAAACTCGCAGAGCAAGAAACCAAAAAACA 3224
913 ..... 913
```

```

3225 GGGGAAAAAGACACGCGCAAGCCTTGAAGCGCTGATTGCGCGCGGCG 3274
913 ..... 913
3275 GCGATGCCCTCGAAAGACAGAAAGCTTCCGACCGCGCGCGCGAGCA 3324
914 ..... 922
3325 GCGGGAATAATGCGGCATTATGACGCGAGAGAAAGAAAAAGCGGT 3374
922 ..... 922
3375 GCAGCGGATPAAAGACACCGCTTGGGAGAAACGCGGAAGCGGAACCC 3424
923 ..ProProAspProAsp..... 927
3425 GCGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGGATTTG 3474
928 ..ProThrProAsp..... 931
3475 CCGCACTGCAACCCCAACGCGACCCCAACGCGAGCGACCTGATCAG 3524
932 ProAspProThrProAspProAspProThrProAspProGluProThrPr 948
3525 CCGTTATGCCAATAGCGGTTGAGTGAATTTCCGCGCACGCTCAACAGCG 3574
948 oAlaTyrGlnProValLeuAsnAlaLysValGlyTyrLeuAsnAsnL 965
3575 TTTTCGCGGTACAGAGCAATTAAGCGCGCTATTTCGCAAGACCGCGCG 3624
965 euAlaGlaAlaL.....AsnGlnAlaPheMetMetGluArgArg 977
3625 AACGCCGTTTGGACAGCGCG.....ATCCGGGA 3653
978 AspnAlaIaGlyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyG 994
3654 CACCAAAACACGCTTTCGCAAGATTTCCGCGCTACCGCGCAACAAACCG 3703
994 yAspTyrHisTyrThrAlaIaGlyGlnLeuAla.....GlnHisGluA 1009
3704 ACCTGCGCAATCGGTATGCAAGAAACCTCGCGAGCGGCGCG..... 3747
1009 spThrSerThrValGlnLeuSerGlyAspLeuPheSerGlyArgTyrPly 1025
3748 .....GTCCGATCCG.....TTTGGCAACACG 3773
1026 ThrAspGlyGlyTyrMetLeuGlyTyrLeuAlaGlyTyrSerAspAsnG 1042
3774 GACGAAACACCTTCGACGAGCGCATCGCACTCGGACGCGCTTGC 3823
1042 nGlyAspSerArgSerAsnMetThrGlyThrArgAlaAspAsnGlnAsn 1059
3824 ACGCGCGCGCTTTGCGG.....CAATACGGCATCGCAC 3855
1059 IsGlyTyrAlaValGlyLeuThrSerSerTyrPheGlnHisGlyAsnGln 1075
3856 AGGTCTCATACGCGCATCAGCGCGCGCGGCTTTAGC.....AGCG 3899
1076 LysGlnGlyAlaTyrPheAspSerTyrPheGlnTyrAlaTyrPheSerAs 1092
3900 CAGCCTTCAGACGCGCATCGAGGCAAAATCCGCGCGCTGCTGCAAT 3949
1092 nAspValSerGlnGlnGlyAspGly.....ThrAspHisT 1104
3950 AC.....GCGATTCAGGACACGATACCGCGCGGCTTC..... 3981
1104 yHisSerSerGlyIleIleAlaSerLeuGlnIaGlyTyrGlnTyrPhe 1120
3982 .....GCGGATTCGCGCATCGGACGCGCATCGGCGCAACGCGCTATT 4025
1121 ProGlyArgGlyValIleGlnProGlnAlaGlnValIleTyrGlnG 1137

```

```

4026 GTCGAAAAAGCGATTACCGCTACGAAAC..... 4056
1137 yValGlnGlnAspPheThrAlaAlaAsnArgAlaArgValSerGln 1154
4057 .....GTCAATTCGCAACCCCGCGCTTGATTCACCGCTACCGC 4098
1154 erGlnGlyAspAspIleGlnThr.....Arg 1162
4099 GCGGCATTAAAGCAGATTATTCATTCAACCGCGCGCAACATTTCCAT 4148
1163 LeuGlyLeuHisSerGluTyr.....ArgThrAlaValHis.....Va 1175
4149 CACGCTTATTGAGCTGTCC...TATACGATCCGCTTCGCGCAAG 4195
1175 IleProThrLeuAsnLeuAsnTyrTyrHisAspProHisSerThrGlu 1192
4196 TCCGAACACGCGCTCAATACCGCGCTATTGGCTCAGGATTTCCGCAAC 4245
1192 LeuGlnGlyAspGlySerThr.....IleSerAspAspAlaValSerGln 1206
4246 CGC...AGTGGCAATGGGCGCTAAACGCGCAATCAAGGTTTCACGCT 4292
1207 ArgGlyGlnIleIleValGlyValThrGlyAsnIleSerGln...ArgVa 1222
4293 GTCCCTTCACGCTGCCCGCGC.....AAAGCCCGCACTGGAAGCGC 4336
1222 IserLeuArgGlySerValAlaTyrPheGlnGlySerAspAspPheAlaG 1239
4337 AACACGCGCGGCGCATCAATTAAGCTACCGCTGG 4371
1239 InThrAlaGlyPheLeuSerMetThrValIleTyr 1250

seq_name: p1r2:H98323

seq_documentation_block:
Hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: H98323
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H98323
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1341 <XOR>
A:Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:gl5160106; GSPDB:GN00170
A:Gene: AGR_L_3085
A:Map position: linear chromosome

alignment_scores:
Quality: 278.00 Length: 1423
Ratio: 0.470 Gaps: 68
Percent Similarity: 41.532 Percent Identity: 20.239

alignment_block:
US-09-303-518D-649 x H98323 ..
Align seq 1/1 to: H98323 from: 1 to: 1341

688 GGCATATCCTTTGCACAAATGATCAGTGGTGGCACATCACTATTAGG 737
1111: ..... 1111
128 GAGTATGAAATTAACATAGCCCATATGCTTTTAAACACAGAGGCGT 787
738 TAGTGAATAATTAACATAGCCCATATGCTTTTAAACACAGAGGCGT 787
144 yAla.....ThrGlyLys 149
788 CATTTGGCAGACAGTGGCTACCAATGTTTATCTATGATGCCCAAGCAA 837

```

149 er.....SerGlyThrAlaSerPheIle.....GlyAla 158
149 er.....SerGlyThrAlaSerPheIle.....GlyAla 158
838 AAGTGTAAATTAATGGGATTAATGCAACGGCC.....870
159 AsnTrpAlaAsnAspGlyAspValSerValGlyGlyAsnGlyThrGlySer 175
871AACCCCTAATAG 883
175 rLeuLeuLeuHisGlyThrThrAlaSerSerGluAsnIleTyrValG 192
884 GAAAGAGC...AATGCTCCAGCTGTTGTAACAT.....918
192 LysIleSerGlyThrGlySerGlySerLeuLysLeuAspAsnSerThr 208
919 ...TGTTCTATGATGAAATCTTGTGGAGATACCAT.....954
209 LeuThrValAlaAspThrValPheAlaGlyTyrAsnSerTrpSerSer 225
955TCAGTATCT 964
225 rAlaAlaGlyGlySerGlyAsnValGluValLeuGlyAlaSerSerLeu 242
965 ACGACACAGCTCAAAATGGGAATFACTCTTTACAGCAGTAAATAGGC 1014
242 ySer.....AlaAsnGlyValLeuAlaAsnAspProAspThrValGly 256
1015 ACAGGAAATCAATGCAAAACATGAACAAATCTCTGCTTAATAGAT 1064
257 ThrAlaLeuValSerGlyGlySerSerTrpAsnMetSerSerSple 273
1065 A.....AAAACAGAACGTTCAATGTTAATG 1093
273 uValValGlyGluAlaGlyValGlyThrLeuThrIleThrGlyGlyGlyM 290
1094 TTCTTTATCCAGACAGAGAACGCTTTATCATGCTGCAGGTGT 1143
290 eValIleSerGlnThr.....295
1144 GTCAAGTTATTCAGACCACTGATTAATGAGAAATATTTCTTTAT 1193
296 ...ThrThrValIleAlaAspLeuAsnSerAlaAspGlySerSerValTh 311
1194 TGACGAGAGAAAGGCGATGATGACTTACAGCAACATC.....A 1234
311 rValSerAlaGlyLeuGlySerSerLeuGlySerSerLeuValAlaGly 328
1235 ATCAAGTCTGAGAGATTAATTCAGAGAGATTTACGGTCCGGCT 1284
328 snGlyValAlaAlaLysLeuValAlaGluAlaAlaIleThrValThr... 343
1285 GAAATTAACGAACCTTGGCAAGC.....GCGGGCGTTCAATATCAG 1325
344SerGlyLeuAlaIleIleGlyArgHisSerAl 354
1326 TGAAGACAGTACCTTACT.....TGG.....1347
354 aserGluAlaThrValThrGlyAspGlySerLysTrpThrThrGlyAspL 371
1348 ...AAAGTAAACGGCGTGGCAACGACCGCTGCCAAATCGGCAAGGC 1395
371 euGluValGlyGlyAspThrSerAspProGlyGlyLeuAlaGlyAsnGly 387
1396 ACGTGCAGTTCAGCAAGGGGA.....AACCA 1427
388 ThrLeuAsnValThrAlaGlyLysSerValAspSerThrValAlaHisIle 404
1428 AGGCTGCATCAGCGTGGCGAGCATGATTTGGATCAGCAGGAG 1477
404 uGlyValValAlaGlyAlaThrGlySerAlaIleValasp.....417
1478 ACGATAAAGCGAAAAACAGCCTTTAGT.....CAA 1509
1478 ACGATAAAGCGAAAAACAGCCTTTAGT.....CAA 1509
418GlyLysGlySerValTrpThrValAlaAspArgAsnSerLeuGlu 431
1510 ATCGGCTGTGACGCGAGGAGGTACGAGTACGTAACGTAATGCCATATCA 1559
432 ValGly...ValSerGlyAlaGlySerLeuAlaValThr.....443
1560 GTTCAACCCCGAACAACTATTTGGCTTTGCCGGCAGCTTTGATT 1609
444GlyGlyGlyLeu...447
1610 TAAAGCGCATTCGCTTCTGTTCCACCGTATTCAAATAACGATGAAGG 1659
448ValAspAlaAla 451
1660 GCGATGATTCACACCAATCAAGACAAGAATCCAGCTTACCATYAC 1709
452 AsnIleIleIleGlyThrAsnThrGly.....460
1710 AGCAATTAAGATATTGCTACACCGGCAATTAACAGCTTTGATGCA 1759
461 .GlyAsnGlySerValArgValSerGlyAlaAspSerThrValLysSer 477
1760 AAAAGAAATTCCTCAACAACGCTTGTGGCGAGAGAATACAGCAAA 1809
477 rGSerAspLeuAsnVal...GlyLeuTyrGly.....486
1810 ACGAAGCGGCGCTCAACCTGTTTACAGCCCGCGCAAGACCGC... 1857
487 ...AsnGlySerMetThrValGluAlaGlyAlaValLysSerArgAs 502
18571857
502 pGlyTyrValAlaThrTyrGlyGlySerThrSerAlaValThrValThrg 519
1858ACCTGCTGCTTCCGCGGA 1878
519 LysAspGlySerSerTrpAlaMetThrGlyThrPhePheValGlyTyrAla 535
1879 ACAATTTAAACGGCAACATACCGCAACAAGCGAACTGTTTTCAG 1928
536 SerGlyAlaThrGlyAsnValThrValSerAsnGlyAlaIle...550
1929 CGGAGACCAACACCGCACCGCTTCAATTCATTAAGACCATGTGTCG 1978
551A 551
1979 AAAAAGAGCGATTCCTCGCGGGAATCGTGGG.....2013
551 rGAlaThrGlyValThrLeuGlyAspLeuAlaGlyAlaSerGlyThrMet 567
2014GACACGACTG 2024
568 ThrIleThrGlyAlaGlySerLysValThrAlaTyrValAspAsnGlyTh 584
2025 GATCAACCGCACATTTAAACGGAACCTTCAATTAAGGCGGACAG 2074
584 rValAsnSerGlySerValAspValGlyPheGlnGlySerGlySerLeuS 601
2075 CGTGGTTCCTCCGCAAT.....GTTGCCAAA 2100
601 eValValAlaAsnGlyGlySerLeuAspAlaTyrAsnLeuTyrValGlyAsn 617
2101 GTGAAGCGCATTCGCTTACCAATCAACGCAAGCAATTTTGGTGT 2150
618 AlaLeuGly.....SerSerGlyAlaValLeuValSerGlyVal 630
2151 CGCACCGCATCAAC.....2166
630 rGlySerHisValSerValAspGlyLeuMetValValGlyAsnAlaGly 647
2167CAC 2169
647 snGlySerValGluIleThrGlyGlyAlaSerLeuAlaIleProThrIle 663

2170 ACAATCTGACAGCTTGCAGCGGCTG..... 2202
2203ACAATCTGTCGAA... AAAACCTTA 2227
680 ySerGlyIThrAlaArgSerAlaGlyAlaValAlaArgAlaLeu 697
2228 CC.....GACGATAAAGTATGCTTCATTGACTAGAAC... 2262
697 IapheGlyAlaGlyAsnGlySerIleValPheAsnHisSerGluThrGly 713
2263GACATCAGCGCAATGTCATCTTCCCATCAGC 2297
714 TyrThrLeuSerAlaAspIleSerGlyAlaGlyAlaValAlaGlu 730
2298 TCATTTAAATCTCAGAGGCTTGCACACTCAAGCAATCTTATGCA 2347
730 a.....GlyValThrThrLeuSerGlyAsnAsnSerTyr 742
2348 ATGGGATACACGTTATACA.....GTACGCCACACGCCACCA 2388
742 erGlyGlyThrThrIleSerAlaGlyMetLeuLysGlyThrAlaLys 758
2389 AACGGCAACCTTAGCTCGTGGGCAATGCCAACACATTATATCAAGC 2438
759 PheGlySerGlyGlyIleValAsnAsnAlaGlyLeuValAlaAsp 773
2439 CACATTAAAGCGCAACACA.....TCGCTTCGGGCAATCTTCAT 2479
774GlyGlyGlyThrLeuSerAsnAlaIleSerGlyThrGlySerP 788
2480 TTAATCTAAGGACCAAGCGCTGACAAAAGCAGCTGACGCTTCCGCG 2529
788 heGlyLysThrGly.....AspGlyAsnLeuLeuThrGly 800
2530 AACGCTAAGGCAAC.....GTAGCCATTCGGCAGCTCAACGG 2567
801 AsnSerThrTyrSerGlyAlaThrAlaValSerAlaGlyLysLeuSerVa 817
2568 TAATGTCTCCCTAGCCGATTAAGCAGTATTCATTTGAAAAGCAGCGCT 2617
817 LAsnGlySerLeuAlaSerAlaVal.....SerValG 828
2618 TTACGGACAATTCAGCGCGGCAAGATAGGCACTTACTTAAA... 2664
828 LysSerGlyAlaThrValGlyGlyThrGlyThrIleGlyGlyLeuThrVal 844
2665 GACAGGATGAGCGCTGCCGTCAGGACGAGATTAAGCAATTAA...AA 2711
845 AsnSerGlyGlyThrLeuAlaProGlyAsnSerIleGlyThrLeuThrSe 861
2712 CCTTGACAAGCGCCACATTACATCAATTCGCGCTAT.....CGCCACG 2755
861 rThrGlyAsnAlaThrPheAlaSerGlySerThrTyrAlaValGluIle 878
2756 ATGCGGCAAGG.....GCCAAGCGGAGTGGCAGCAAGT 2790
878 sPAlaSpIleYerSerAspArgLeuAlaValThrGlyThrIleThrIle 894
2791 GGGCGCGCGCGCGTTCGCGCGCTTCCTTATATCCGTTAC 2840
895 Ala.....AsnAspValSerLeuIleValIth 903
2841 ACCG.....CSAATTCGGTGAATCCGTTTCAACAGCG 2875
903 rProLeuGlyAlaHisSerAlaTyrSerLeuGlyThrArgTyrThrIle 920
2876 TGAAGTAAAGCAATGAAAGCGTCAAGCAATTCGCTTAATGTCG 2925
920 eThrAlaThrGlyGlyValThr.....GlyThrPheSerValAsp 934

2926 GAATCTTCGGCTACCGGACGACAAATTAAGCTGGCGGAAGTTCCGA 2975
935 GluSerPheAlaTyrLeuThrAlaLysValAlaGlnSerGlyAspAspAl 951
2976 AGGCATTCAC.....ACCT 2989
951 eThrThrTyrLeuSerPheLeuArgThrSerProAspSerGlyLeuLeu 968
2990 TGGCGGTCAACAATACCGGCACGAACTGCAAGCTCGCAACATTCAGC 3039
968 lAlaAlaThrSerThrAlaAsnAlaArgAlaAlaAlaValAlu 984
3040 GTAGTGAAGAAAAGACAACAACCGCTGTCGAAAACCTTAATTTCAC 3089
985 AlaLeu.....GlyGlnThrSerProLeuTyrGluAlaAlaLeuPheLe 999
3090 CTGCAAAAACGACAGCTGATGCCGGCGCGGTGATCAACATTCATCC 3139
999 uGlnGlnGlyGlu.....ThrGlnGlyAlaPheSer...GlnLeuAla 1012
3140 GCMAAGCGCGAGTCCGCTGCATTAATCCGTCMAAGAACAGAGCTT 3189
1013GlyLeuIle.....HisProSerLeuAlaMetAlaLeu 1023
3190 TCCGACAAACTCGGCAGACAGACCAAAAAAACGCGGAAAAAGACA 3239
1024 lLeAsnArgSerGly..... 1028
3240 CGGCAAAAGCTTAGCGCGCTGATTCGGCGCGGCGCGATGCGTGANA 3289
1029GlnSerArgAspValAlleuAsnArgLeuArgSerAlaPheGlu 1044
3290 AGACAGAAAGCGTTGCCAAACCGCGCGGACGAGCAGCGGGGAAAAATGTC 3339
1044 LyAlaAsp..... 1046
3340 GGCATTATGACGCGCGAGAGAGAAAGAGGGTCCAGCGGATTAAGA 3389
1046 1046
3390 CACCGCTTGGCAAAACAGCGGACGGAACCGCGCGCTACACAGC 3439
1046 1046
3440 CTTTCCCGCGCGCGCGCGCGCGGATTTGCCGCACTGCACACC 3489
1047AlaArgProIleLeuPro..... 1052
3490 CAACCGACGCCCAACCGCAGCGGACGTGATCAGCGCTTATGCAATAG 3539
1053ValAlaTyrAlaGlu 1058
3540 CGGTTAGTGAATTTCCGCCACGCTCAACAGCGTTTTCGCGTACAG 3589
1058 yGlyGlnAsnProPheAla..... 1064
3590 ACGAATTAGACCGGTATTTCCGAGAGCCGCGCAACGCCGTTTGACA 3639
1065ValAspAspGlyAlaLeuSerPheTyrPser 1074
3640 AGCGGCAATCCGGGACACCAAAACTACCGTTCCGAAATTTCCGCGCTA 3689
1075 SerGlyPheArgSer..... 1079
3690 CCGCAACAAACGACCTCGGCCAATCGGTATGCAAAAAACCTCGCA 3739
1080 ArgGlyArgIleAspSerAspGlyAsnGlySerValAspMetAsnG 1096
3740 GCGGCGCGTGCATCTGTTTTCGACAAACGAGCAACCAACACCTTC 3789
1096 LysGly.....GlyIleLeuPheGly.....Leu 1103
3790 GACGACGCACTCGGCACTCGGCAAGCGCTTGCCACGCGCGCTTTCGG 3839

```

1104 AspGlyGluLeuSerAspSerTrpArgLaglyValAlaIaLaglyTrpG 1120
3840 GCATACGGCATCGACAGGTTTACATCGGCATACGGCGGGGGCGGGT 3889
1120 YHLSAspGlyIleLysGlnThrAlaLeuAlaAlaSerAla..... 1133
3890 TTAGCAGCGGCGACCTTTTCAGACGGCATCGAGCAAAATCCGCGCGCG 3939
1133 ..... 1133.
3940 GTGCTGCTATTACGGCATTTACGACGATACCGCGCGGTTCGCGCAT 3989
1134 .....AspValAspSerTrpTrpLeuAlaIaLaglyTrpG 1146
3990 CGGCATCGAACCG.....CACATCGCGCGCAACCGCGCTTTTCGTC 4030
1146 rGlnIleGlyProAlaSerLeuArgLeuGlyAlaIleHisAlaPhe...G 1162
4031 AAAAAGCGATTACCGCTACGAAAAAGTCATATCGCCACCCCGCGCTT 4080
1162 LAspAlaLaglyThrArg.....ArgSerIle 1170
4081 GCATTACACCGGTACCGCGCGCATTAAGCATATTATTCATAAAC 4130.
1171 SerPheSerThrLeuGlnIleuSerLeuAlaSerTrpGlyGlySerTh 1187
4131 GCGCAACACATTTCC..... 4146
1187 rSerGlnValPheAlaGluAlaIaLrPhePheAspPheSerLeuThr 1204
4147 ..ATCAGCGCTTATTGAGCTGTCTCTTACCGATCGCGCTTCGGGCA 4194
1204 IsIleGluProTrpAlaAsnIleAlaIaTrp..... 1213
4195 GTCCGACACGCGTCAT.....ACCGCGTATTGGCTCA 4229
1214 ValAsnThrArgThrAspGlyPheGlnIleuLysGlyAlaIleAlaIa 1230
4230 GGATTTGGCAAAACCGCG 4248
1230 LserSerGlySerThrArg 1236
seq_name: p1r2:S28634
seq_documentation_block:
adhesin AIDA-I precursor - Escherichia coli plasmid PIB6
C:Species: Escherichia coli
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28634; S22680; S28881; S72657
R:Benz, I.
submitted to the EMBL Data Library, March 1992
A:Reference number: S28634
A:Accession: S28634
A:Molecule type: DNA
A:Residues: 1-1286 <BEN>
A:Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA46156.1; PID:g42255
R:Benz, I.; Schmidt, M.A.
Mol. Microbiol. 6, 1539-1546, 1992
A:Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escheri
A:Reference number: S22680; MOID:92326638
A:Accession: S22680
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 839-1286 <BE2>
A:Cross-references: EMBL:X65022
A:Experimental source: strain 2787
A:Accession: S28881
A:Molecule type: protein
A:Residues: 50-56 <BE3>
A:Experimental source: strain 2787
R:Suhr, M.; Benz, I.; Schmidt, M.A.
Mol. Microbiol. 22, 31-42, 1996

```

```

A:Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the out
A:Reference number: S72657; MOID:97055419
A:Accession: S72657
A:Molecule type: protein
A:Residues: 847-856 <SUN>
A:Experimental source: DAEC strain 2787
A:Genetics:
C:Keywords: membrane protein
C:Keywords: signal sequence #status predicted <Sig>
F:1-49/Domain: adhesin AIDA-I #status predicted <Mat>
F:50-1286/Product: adhesin AIDA-I #status predicted <Mat>

alignment_scores:
Quality: 275.00 Length: 1458
Ratio: 0.434 Gaps: 65
Percent Similarity: 43.416 Percent Identity: 18.999

alignment_block:
US-09-303-518D-649 x S28634 ..

Align seg 1/1 to: S28634 from: 1 to: 1286

688 GCATACCTTTGCAAAATGATCAGT.....GGTGGCAC 725
|||||:||||| ||| |||||
45 GlyAsnAlaPheAlaValAsnIleSerGlyThrValSerSerGlyGlyTh 61
726 AGTCACTAGTAGTAAATAAATTAACATAGCCATATGT..... 768
|||||:||||| ||| |||||
61 rValSerSerGlyGlnThrGlnIleValTrpSerGlyArgGlyAsnSer 78
769 .....TTTTACCAACAGAGGC... 786
78 snAlaThrValAsnSerGlyGlyThrGlnIleValAsnAsnGlyGlyLys 94
787 .....TCATTGGCGACAG 800
95 ThrThrAlaThrThrValAsnSerSerGlySerGlnAsnValGlyThrse 111
801 TGGCTCAACAATG.....TTTATCTATGATGCCCAAAACCAAGTGT 844
|||||:||||| ||| |||||
111 rGlyAlaThrIleSerThrIleValAsnSerGlyGlyIleGlnArgVal 128
845 TAAATTAAGGATTTGCAACGCGCAACCCCTATATAGCAAAACCAAT 894
|||||:||||| ||| |||||
128 rSerGlyGlyValAlaSerAlaThrAsnLeuSerGlyGlyAlaGlnAsn 144
895 GCGTTCAGCTGTGCTGTAAGATGGTCTATGATGAATCTTGTCTG 944
|||||:||||| ||| |||||
145 IleTrpAsnLeuGlnIleHisAlaSer.....AsnThrValIlePheSer 159
945 AGATACCATTCAGTATTC..... 963
159 yGlyAsnGlnThrIlePheSerGlyGlyIleThrAspSerThrAsnIleS 176
964 .....TACGACCAACGTCAAATGCAAAATCTTTTAACGAC 1002
176 rSerGlyGlyGlnIleArgValSerSerGlyGlyValAlaIaSerAsnThr 192
1003 GATATTAATGCGACAGAAATCAATGCCAACATGAACACAAATCTCT 1052
|||||:||||| ||| |||||
193 ThrIleAsnSerSerGlyAlaGlnAsnIleLeuSerGlyGlyAlaI 209
1053 GCCTAATAGATTAAACACGACGATTCATTTGTTATGTTCTTTAT 1102
|||||:||||| ||| |||||
209 rSerThrHisIleSerSerGlyGlyAsnGlnTrp.....IleSerAlaG 224
1103 CGAGACACCAAGACACTGTTATCATGTCGACGGTGTGTCACAGT 1152
|||||:||||| ||| |||||
224 LValAsnAlaThrGlnThrIleValAsnSerGlyGly..... 236
1153 TATCGACCGAGCTGATTAATGGA..... 1176
|||||:||||| ||| |||||

```

```

237 ... PheGlnArgValAsnSerGlyAlaValAlaThrGlyThrValLeuSe 252
1177 .....GAAMATATTTCTTTATTGACAGAGAAAAAGCGAATTGA 1216
252 rGlyGlyThrGlnAsnValSer.....SerGlyGlySerAlaI 265
1217 TACTTACGACACATCATCAAGTGTGCGAGGATTATATTTCACGA 1266
265 IeSerThrSerValTyrAsnSerGlyValGlnThrVal...PheAlaGly 280
1267 .....GATTTACGCTGTCTGCTGAAATATACGAAACTTGCA 1304
281 AlaThrValThrAspThrThrValAsnSerGlyGlyAsnGlnAsnIleSe 297
1305 AGCGGGGGGGCTTCATATCATGACAGACAGTACCGTTACTTGGAAAGTAA 1354
297 rSerGlyGlyIle...ValSerGlu.....ThrThrValAsnValS 310
1355 ACGGCGGCGCAAAC.....GACCGCTGTCCAAATATCCGCG 1389
310 eRgLyThrGlnAsnIleTyrSerGlyGlySerAlaLeuSerAlaAsnIle 326
1390 AAAGGACAGCTGCACGTTCAAGCCAAAGGGAA..... 1422
327 LysGlySerGlnIleValAsnSerGlyGlyThrAlaIleAsnThrLeuVa 343
1423 .....AACCAAGCTGCATCAGCTGGCGG 1447
343 lSerAspGlyGlyTyrGlnHisIleArgAsnGlyGlyIleAlaSerGly 359
1448 ACGGTACAGCATTTTGTGATCAGCAGCAGCATTAAGCAAAAAACAA 1497
360 .....ThrIleValAsnGlnSerGlyTyrValAsnIleSerSerGly 373
1498 GCCTTATGATGAATCGCTGTGTCAGCGGACAGGCTGACACTGAA 1547
374 GlyTyrIleGlnSerThrIleIleAsnSerGlyGlyThrLeuArgValIle 390
1548 TGGCGATATACGTTCAACCCGCAACACTATTTCGGCTTCGGCGG 1597
390 uSerAspGlyTyrAlaIleArgGlyThrIleLeuAsn.....AsnSerG 404
1598 GACGTTTGATTTAAACGGCGCATCGCTTCGTTCCACCGTATTCAAAT 1647
404 lYArgGlnAsnValSerAsnGlyValSerTyrAsnAlaIleAsn 420
1648 ACCGATCAAGGGCGGATGATTGTCAACCAATCAAGCAAAAGATCCAC 1697
421 ThrGlyGlyAsnGlnTyrIleTyrSerAspGlyGlnAlaThrAlaAlaI 437
1698 CGTTACCATTCAGCAGCATTAAGATATTGTACAAACGGCAAT..... 1740
437 eValAsnThrSerGlyPheGlnArgIleAsnSerGlyGlyThrAlaProv 454
1741 ..AACACAGCTTGATAGCAAAAAAGAAATTGCC..... 1773
454 aGlnAsnSerValValValThrArgThrValSerSerAlaAlaLysPro 470
1774 .....TACAAGGT..... 1782
471 PheAspAlaGlnValTyrSerGlyGlyLysGlnThrValTyrLeuThrPar 487
1783 .....TGTTTGGCGAGAAAGATACG..... 1803
487 gGlyIleThrPtyrSerAsnPheLeuThrAlaValTrpSerMetPhePro 504
1804 .....ACCAAAACGACGGCGGCTCAAC..... 1827
504 lYThrAlaSerGlyAlaAsnValAsnLeuSerGlyArgLeuAsnAlaPhe 520
1828 .....CTTGT 1832
521 AlaGlyAsnValValGlyThrIleLeuAsnGlnGlnGlyArgGlnTyrVa 537
1833 TTACAGCGCCGCGCA.....GAAGACCGCA 1858
537 lTyrSerGlyAlaThrAlaThrSerThrValGlyAsnAsnGlnGlyArgG 554
1859 CCGTCGCTGCTTCCGCGGAAACAATTTAAACGGCAATCAGCAACACA 1908
554 lutyValLeuSerGlyGlyIleThr...AspGlyThrValLeuAsnSer 569
1909 AACGGCAACAGCTTTTCAGC.....GGCAGACCAACACCGCAGGCTA 1952
570 GlyGlyLeuGlnAlaValSerSerGlyGlyLysAlaSerAlaThrValI 586
1953 CAATCATTTAAACGACCATTTGTGTGCAAAAAGAGGCAATTCCTCGCGGG 2002
586 eAsn.....GlnGly.....GlyAlaG 592
2003 AAATCGTGTGGACACAGCAGTGCATCAACCCGACATTTAAAGCGGAAC 2052
592 lnPheValTyrAspGlyGlyGlnValThrGlyThrAsnIleLysAsnGly 608
2053 TTCCAAATTTAAAGCGGACAGCGGCTGTCCGCAATTTGGC...AA 2099
609 GlyThrIleArgValAspSerGlyAlaSerAlaLeuAsnIleAlaLeuSe 625
2100 AGTGAAGCGCATTTGCAATTTGACCAATCACGCC..... 2133
625 rSerGlyGlyAsnLeuPheThrSerThrGlyAlaThrLeuProGluLeu 642
2134 .....CAACGATTTTGTGTGCGCACCG...CAACCAAGCACACATC 2175
642 hThrMetAlaIleAsnSerValSerGlnAsnHisAlaSerAsnIleVal 658
2176 TGTACACGTTGCGACTGCGAGCGGTCTGCAAAATTTGTGCAAAAACAT 2225
659 LeuGlnAsnGlyGlyLeuLeuArgValThrSer.....GlyGlyThrAl 673
2226 TACCGACGATTAAGATGATTTCTTCAATTGACTAAGCCACATCAGCGCA 2275
673 aThrAsp.....ThrThrValAsnSerAlaGlyA 683
2276 ANTCGATGTCGCGATCAGCTCATTTAAATTCACAGCGCTTCGCCCA 2325
683 rGleuArgIleAspAspGlyGlyThrIleAsn.....GlyThrThr 697
2326 CTCACAGCC..... 2334
698 lLeAsnAlaAspGlyIleValAlaGlyThrAsnIleGlnAsnAspGlyAs 714
2335 .....ATCTTAGTCMAATGGCGATACAGTTATACAGTCAGCC 2374
714 nPheIleLeuAsnLeuAlaGlnAsnTyrAspPheGlnThrGluLeuSerG 731
2375 ACACAGCC.....ACCCAAAACGGC 2394
731 lYserGlyValLeuValLysAspAsnThrGlyIleLeuThrThrArgIaGly 747
2395 AACCTTAGCCTC.....GTGGCAATGCCCAACCAAT 2429
748 ThrLeuThrGlnAlaGlnGlyValAsnValLysAsnGlyIleIleLeh 764
2430 TATTCAGCGCACATTAACGGCAACATCGCTTCGGCAATGCTTCAT 2479
764 eAspSerAlaValValAsnAlaAspMetAlaValAsnGlnAsnAlaTyrI 781
2480 TTAATCTAACGACCAAGCGCGATCAAAACGCAAGTGTACGCTTCGCGG 2529
781 lAsnIleSerAspGlnAlaThrIleAsnGlySer.....ValAsnAsn 795
2530 AACGCTAAGCAACAGTAAAGCATTCGCGACCTCAAGGTAATGATCCCT 2579
796 AsnGlySerIleValIleAsnAsnSerIleIleAsnGlyAsnIleThr.. 811

```


917 alAlaglySerLeuGluSerLeuSer...LeuThrGlyLysIrrgAlaasp 932
2950 AAATTGAAGTGGCGGAAAGTTCGGAAGGACCTTACACTTGGCGGTCA 2995
933 PheAlaasnSerAlaLysAlaSer...GlyLeuLeuAspLLeSerAlaas 948
3000 CAATACCGGCAACGAACCTGCAGACCTCGACAAATTGACGGTAGCGAG 3049
948 pser.....GlnIleIleMetAsnArg 956
3050 GAAAGAGCAACAACCGCTGTCCGAAACCTTAATTTCACCTGCAAAAC 3095
956 LysAlaaspThrAlaGlnAlaAsnLeuAsnLeu..... 966
3100 GAACACGTCGATGCCGCGCGTGGCTTACCAACTCATCCGCAAGAC... 3147
967AlaGlyArgLeuGluLeuLeuAlaSerAspVa 977
3148GGCGAGTTCCGCCTGCATTA 3166
977 LAlaGlnAlaValAlaGlnProValAlaArgAlaAlaMetGluLeuSerA 994
3167 AT..... 3168
994 snAlaIrrgAlaValMetProAlaProAlaMetProValProAlaAla 1010
3169 CCGGTCAAGAACACAGAGCTTTCGACAAATCGGCACAGGCAAGCCAA 3218
1011 ProValAlaGlnIlePheAlaLeuAsnAspValValMetThrGlyGlyThrVa 1027
3219 AAACACGCGGGAAGAAAGACACGCGCA...AGCCTGACGGCGTATG 3265
1027 LaspMetSerAsnAlaLysAsnAlaGlnLeuThrMetLAserLeuAsnG 1044
3266 CGCGCGGCGGC...GATGCCGTGCAAAACACAGAAAGCGTTGCC 3306
1044 LThrGlyAsnPheAsnLeuGlySerValMetGlnIrrgAspSerValAla 1060
3307 GAACCGCGCGGCAAGGCGGGGAATAATGCGGCATTATGACGGCGGA 3356
1061 ...ProLeuAsnValSerLLeLysAlaAsnGlyAspPheIleIleAlaMe 1076
3357 GGAAGAGAAAAACGGGTGACAGCGGATTAAGACCGCTTGGCGAAAC 3406
1076 LAsnSerSerGly.....GlnAlaProThrAsnLeuAsnValValAsnT 1091
3407 AGCGGGAACGGGAACCCGCGCGGTACACCGCTTCCC..... 3447
1091 hrAsnGlyLysAlaIrrgPheAlaLeuAlaAsnGlyProValAlaLeu 1107
3448CGCGCGCGCGCGCGCGCGGCGGATTTGCGCGCA 3479
1108 GlyAsnTyrrMetThrAsnLeuAlaLysAspAlaAsnGlyAsnPheValLe 1124
3480 ACTGCAACCCCAACCGCAGCCCAACGCGACGCGACCTGATCAGCGGT 3529
1124 uThrAlaAspLysSerAlaMetThrProGlyThrAlaGlyLeuAlaIrr 1141
3530 ATGCAAAATGACGGTTGAGTGAATTTTCCGCGACGCAACAGCGTTTC 3579
1141 alAlaAsnThrThrProValIlePheAsnAlaGluLeuSerSer..... 1155
3580 GCCGTGACGAGCAAGATTAGACCGCGGTATTGGCGAAGACCGCGCAAGC 3629
1156 ...LleGlnGlnIrrgLeuAspLysGlySerThrGlyuThrAsnGlnSerG 1171
3630 CGTTTGGACAGCGCGCATCGCGGACACCAAGACTACCGTTCCAGATT 3679
1171 TrrrGlyIrrgLeuSer.....TyrrLeuAsnAsnAsn 1181
3680 TCCGCGCTACCGCAACAAACGACCTGCGCAAAATCGGTATGCAAAA 3729

1181 heAlaVal..... 1183
3730 AACCTCGCAGCGGCGCTCGGCATCTGTTTCGACAAACCGGACCGA 3779
:::|||||:::
1184LysGlyArgAlaAla..... 1188
3780 AAACACCTTGACAGCGCATCGCAACTGCGACGGCTTGCCACGGCG 3829
:::|||||:::
1189AsnPheAspGlnLysLeu.....AsnGlyW 1197
3830 CCGTTTCGGGCAATACGCATCGACAGTTTACATCGGCATCAGCGCG 3879
:::|||||:::
1197 etThrLeuGlyGlyAspLysAlaThrAlaLeuAlaAspGlyValLeuSer 1213
3880 GCGCGGGGTTTAAACAGCGGCGAGCTTTCAGAC.....GG 3914
:::|||||:::
1214 ValGlyGlyPheAlaSerTyrSerSerSerAspIleYThrAspTyrG1 1230
3915 CATCGAGGCAAAATCCGCCGCCGCTGTCATTCAGCATTCAGGCAC 3964
1230 nSerLysGlyLysValAspSerHis.....SerPheGlyAlaTyrAlaG 1245
3965 GATACCGCGCGGCTTTCGCGGATTCGCATCGAACCGCACATCGGCGCA 4014
1245 LntYrLeuAlaAsnSerGly..... 1251
4015 ACGCGCTATTTC.....GTCCAAAAGCGGATTACCGGTACGAAA 4055
1252TyrTyrMetAsnAlaValAlaLysAsnAsnGlnPheSerGlnAs 1266
4056 CGTCAATTCGCGCACCCCGGCTTGATTCACACCGCTACCGCGGCGCA 4105
1266 pValAsnIleThr.....SerIleAsnGlySerAlaSerGly 1279
4106 TTAAAG.....GCAGATTATTCATCAAAACCGCGCAACACATT 4143
1279 alSerAsnPheSerGlyMetGlyIleAlaLeuLysAlaGlyLysHisPhe 1295
4144 TCC.....ATCAGCGCTATTGTAGCGCTGTCCTATACCGA 4178
1296 AsnPheAsnGlnAlaTyrValSerProTyrValAlaMetSer.....Al 1310
4179 TGCCGCTTCGGGCAAAAGTCGACACACGCGTCATACCGCGTATTGGCTC 4228
1310 aPheSerSerGlyLysSerAsnIleSerLeuSerAsnGlyMetGlnAlaG 1327
4229 AGGATTTCGGCAAAACCGCAGTGC.....GAATGGGCGGTAAAGCC 4272
1327 InSer...SerSerThrArgSerAlaMetGlyThrLeuGlyValAsnAla 1342
4273 GAAATCAAAAGTTTCACGCTGCTCCACGCTCCGCGCGCAAAAGGCC 4322
1343GlyTyrArgPheValMetAsnAsnGlnAlaGlnLeuLysPr 1356
4323 G 4323
1356 o 1356